Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Scuola di Dottorato in
Scienze Statistiche
Ciclo XX

Dynamic models for competing risks and relative survival

Giuliana Cortese

Direttore: Prof.ssa A. Salvan

Supervisore: Prof.ssa A. Salvan

Co-supervisori: Prof. P. K. Andersen, Prof. T. H. Scheike

29/02/2008

Contents

Summary

The thesis concerns regression models related to the competing risks setting in survival analysis and deals with both the case of known specific causes and the case of unknown (even if present) specific causes of the event of interest.

In the first part, dealing with events whose specific cause is known, competing risks modelling has been applied to a breast cancer study and some of the dynamic aspects such as time-dependent variables are tackled within the context of the application. The aim of the application was to detect an optimal chemotherapy dosage for different typologies of patients with advanced breast cancer in order to control the risk of cardiotoxicity. The attention was concentrated on the cumulative incidence probability of getting cardiotoxicity in a well-defined time period, conditional on risk factors. This probability was estimated as a function of the time-dependent covariate dosage. Within the context of the application, some problems of goodness-of-fit related to time-dependent covariates are discussed.

The previous application gave rise to investigating the role of time-dependent covariates in competing risks regression models. There exist various types of time-dependent covariates, which differ in their random or deterministic development in time. For so-called internal covariates, predictions based on the model are not allowed, or they meet with difficulties. We describe a general overview of the state of the art, problems and future directions. Moreover, a possible extension of the competing risks model, that allows us to include a simple random binary time-dependent variable, in a multi-state framework, is presented. Inclusion of the sojourn time of an individual in a certain state as a time-dependent covariate into the model, is also studied.

In the second part of the thesis, dealing with events whose specific cause is unavailable, regression models for relative survival are discussed. We study the nonparametric additive excess hazards models, where the excess hazard is on additive form. We show how recent developments can be used to make inferential statements about this models,

and especially to test the hypothesis that an excess risk effect is time-varying in contrast to being constant over time. We also show how a semiparametric additive risk model can be considered in the excess risk setting. These two additive models are easy to fit with estimators on explicit form and inference including tests for time-constant effects can be carried out based on a resampling scheme. We analyze a real dataset using different approaches and show the need for more flexible models in relative survival.

Finally, we describe a new suggestion for goodness-of-fit of the additive and proportional models for relative survival, which avoids some disadvantages of recent proposals in the literature. The method consists of statistical and graphical tests based on cumulative martingale residuals and it is illustrated for testing the proportional hazards assumption in the semiparametric proportional excess hazards model.

Riassunto

La tesi riguarda modelli di regressione per rischi concorrenti in analisi di sopravvivenza, e tratta sia il caso in cui le cause specifiche di un evento sono note sia il caso in cui tali cause sono sconosciute, pur se esistenti.

La prima parte della tesi, relativa alle cause specifiche note, presenta un'applicazione del modello di regressione per rischi concorrenti per lo studio sul cancro della mammella. Nell'ambito di questa applicazione, sono affrontati alcuni aspetti dinamici del modello, come per esempio le variabili esplicative dipendenti dal tempo. Lo scopo dell'applicazione è consistito nell'individuare un dosaggio chemioterapico ottimale per diverse tipologie di pazienti con cancro della mammella, al fine di tenere sotto controllo il rischio di cardiotossicità. L'attenzione si è concentrata sulla probabilità d'incidenza cumulata di sviluppare la cardiotossicità in un predeterminato periodo temporale, condizionatamente a determinati fattori di rischio d'interesse. Questa probabilità è stata stimata come una funzione della variabile esplicativa dipendente dal tempo, 'dosaggio'. Alcuni problemi sulla bontà di adattamento del modello, in relazione alle variabili esplicative dipendenti dal tempo, sono discussi nell'ambito dell'applicazione.

La suddetta applicazione ha fornito uno spunto nell'esaminare il ruolo delle variabili dipendenti dal tempo nei modelli di regressione per rischi concorrenti. Esistono diverse tipologie di tali variabili, che si differenziano a seconda del loro andamento casuale o deterministico nel tempo. Nel caso delle cosiddette variabili interne, le previsioni basate sul modello non sono possibili o incontrano delle difficoltà. Nella tesi vengono descritti lo stato dell'arte, i problemi e le future direzioni di ricerca in questo campo. Inoltre, nell'ambito dei modelli multi-stato, viene presentato un'ampliamento del modello per rischi concorrenti che permette di includere al suo interno una variabile casuale binaria dipendente dal tempo. La tesi tratta anche l'inclusione del tempo di permanenza in un certo stato del modello come variabile esplicativa dipendente dal tempo.

La seconda parte della tesi, riguardante eventi le cui cause specifiche son sono disponibili, discute i modelli di regressione per la sopravvivenza relativa. Viene studiato il modello non parametrico per i rischi additivi in eccesso, nel caso in cui anche il rischio in eccesso sia in forma additiva. Viene mostrato come alcuni recenti sviluppi possono essere usati per fare inferenza relativamente a tale modello e, in particolare, per verificare che l'effetto di una certa variabile sul rischio in eccesso sia costante, piuttosto che dipendente dal tempo. La tesi presenta anche un modello semiparametrico per i rischi additivi in eccesso. I suddetti modelli, non parametrico e semiparametrico, hanno stimatori in forma esplicita ed i test d'ipotesi sulla costanza degli effetti nel tempo possono essere basati su uno schema di ricampionamento. Un insieme di dati reali è stato studiato usando diversi modelli statistici al fine di evidenziare la necessità di modelli flessibili nell'ambito della sopravvivenza relativa.

In conclusione, viene discusso un suggerimento per valutare la bontà di adattamento dei modelli per la sopravvivenza relativa. Tale proposta consiste in test statistici e metodi grafici basati sui residui di martingala cumulati, e non presenta alcuni degli svantaggi osservati nei recenti metodi offerti dalla letteratura. La proposta è illustrata tramite la verifica dell'assunzione di proporzionalità dei rischi nell'ambito del modello semiparametrico per i rischi proporzionali in eccesso.

Introduction

The thesis concerns regression models related to the competing risks setting in survival analysis. The work deals both with the case of known specific causes and with the case of unknown (even if present) specific causes of the event of interest. In the first case, we discuss the competing risks model and we focus on regression for the cumulative incidence probability. In the second case, where the event related to a certain group of diseased patients is recorded without any cause, regression models for relative survival are discussed. Along all the work, attention is directed towards inferential problems concerning dynamic aspects of models, such as time-dependent covariates and time-varying regression coefficients.

The thesis consists of four chapters, which we have attempted to make self-contained. For this reason, some basic results are recalled more times in order to be able to read each chapter separately.

Chapter 1 provides some background theory on survival analysis, explained using the martingale theory and counting process representation. The competing risks model is also briefly presented within the framework of multi-state models. While Chapter 1 serves as a general background of the thesis, each of the other chapters is introduced by a specific section, where some of the methods and literature relevant for the research work of the chapter are presented.

Chapter 2 of the thesis deals with the case of events whose specific cause is known. In this context, within the framework of a multi-state approach (Andersen and Keiding, 2002), competing risks models and time-dependent covariates are discussed.

Competing risks modelling has been here applied to a breast cancer study. Some of the dynamic aspects such as time-dependent variables are tackled within the context of the application.

The aim of the application, besides illustrating the available methodology for studying competing risks, was to detect an optimal chemotherapy dosage for different typolo-

gies of patients with advanced breast cancer in order to control the risk of cardiotoxicity. The attention was concentrated on the cumulative incidence probabilities for the cause-specific events in a well-defined time period. In a multi-state approach, the cumulative incidence probability of getting cardiotoxicity, conditional to risk factors, was estimated as a function of the time-dependent covariate dosage. Some problems of goodness-of-fit related to time-dependent covariates are discussed.

The application to breast cancer in Chapter 2 gave rise to investigating the role of time-dependent covariates in competing risks regression models, and more generally, in multi-state regression models. There exists various types of time-dependent covariates, which differ in their random or deterministic development in time. When some of these are studied, predictions based on the model are not allowed, or they meet with difficulties. The outlook in Chapter B describe a general overview of the state of the art, problems and future directions. Moreover, a possible extension of the competing risks model, that allows us to include a simple random binary time-dependent variable, in a multi-state framework, is presented. Inclusion of the sojourn time of an individual in a certain state as a time-dependent covariate into the model, is also studied.

The following chapters of the thesis deal with cases where information on causes of death, remissions, etc. is sometimes unavailable, as typically happens in population-based and clinical observational studies with long follow-up. In some situations, this information is recorded on medical registries but it is incomplete or misleading, because death could be only partially due to the disease of interest and it is difficult to classify deaths due to other causes indirectly correlated with the disease of interest. For this reason, the use of cause-specific survival in the framework of competing risks, where at least two distinct alternative causes need to be specified, is problematic. The relative survival approach provides a solution to these difficulties. It does not require information on cause of death, while it allows to estimate patient survival corrected for the effect of other causes of death, using the natural mortality of the underlying population. Indeed, relative survival describes the excess mortality for patients diagnosed with the disease of interest, irrespective of whether the excess mortality is directly or indirectly attributable to the disease.

In Chapter 3 within the context of relative survival, we study the additive excess hazards models (Zahl, 1996), where the excess hazard is on additive form. We show how recent developments (Scheike, 2002) can be used to make inferential statements about the nonparametric additive excess hazards model. This makes it possible to test the key hypothesis that an excess risk effect is time-varying in contrast to being constant

over time. One problem with the fully nonparametric dynamic description is that the model might be too big, if some covariate effects are in fact constant with time. Therefore, we also show how a semiparametric additive risk model (McKeague and Sasieni, 1994) can be considered in the excess risk setting. This model can provide a better and more useful summary of the data and makes a better bias/variance trade-off. We show how these two additive models are easy to fit with estimators on explicit form and how inference including tests for time-constant effects can be carried out based on a resampling scheme. We analyze a real dataset using different approaches and show the need for more flexible models in relative survival.

A parallel objective of the thesis is to assess the importance of time-varying effects for regression models in the relative survival framework, showing their advantages especially within nonparametric and semiparametric regression models. Presence of time-varying coefficients in the model shows directly how the influence of risk factors on the excess hazard may change over follow-up time. No difficulties appear in handling time-dependent covariates, which are treated as commonly performed in the Aalen additive hazards model and in the Cox model.

There is a general lack of accomplished methodology for the regression diagnostics and assessment of goodness-of-fit of additive relative survival models. The existing theory is only sometimes implemented in public software. In Chapter 4 we describe a new suggestion for goodness-of-fit of the additive and proportional models, which avoids some disadvantages of recent proposals in the literature (Stare et al., 2005). It consists of statistical and graphical tests based on cumulative martingale residuals (Lin et al., 1993). The method is illustrated for testing the proportional hazards assumption in the semiparametric proportional excess model (Sasieni, 1996). This approach is very simple to implement and is known to work well in the standard survival setting. An application based on real data is used to show how these techniques work.

Note: The application to breast cancer presented in Chapter 2 was developed jointly with Marianne Ryberg and Dorte Nielsen, from Herlev Hospital, University of Copenhagen, Denmark, who collected the clinical data from various observational studies.

Chapter 1

Survival Data and Theoretical Background

About twenty years ago there was an extensive development of the theory of statistical models based on the counting process representation in the field of survival analysis. Nowadays, research in this area is mostly based on those developments, which have opened new perspectives on possible extensions and alternative solutions, especially for nonparametric and semiparametric models.

In this chapter, the general framework and some of its basic concepts and methods will be presented, aiming to explain the theory underlying this thesis. The principal references for this chapter are Fleming and Harrington (1993), Andersen et al. (1993), Therneau and Grambsch (2000) and Martinussen and Scheike (2006).

1.1 An overview on survival data

Survival analysis deals with data where the random variable under study is the time T^* from a well-defined time origin to the occurrence of a certain given event of interest. General extensions concern the study of multiple temporal variables or multiple events of interest. The main feature of survival data is the presence of incompletely observed survival times. Incompleteness can be of different type, the most common example being the right censoring, which is explained in detail later.

1.1.1 Survival and hazard functions

We consider a random survival time T^* that has probability density function $f(\cdot)$. The $F(\cdot)$ denotes the distribution function of T^*. The distribution of T^* is often equivalently characterized by the survival function

$$S(t) = 1 - F(t) = P(T^* > t).$$

The hazard function, also called instantaneous rate, is defined as

$$\alpha(t) = \frac{f(t)}{S(t)} = \lim_{h \downarrow 0} P(t \leq T^* < t + h \,|T^* \geq t)/h, \tag{1.1}$$

which represents the instantaneous probability.

The survival function can be computed from the cumulative hazard function $A(t) = \int_0^t \alpha(s)ds$ or equally from the hazard function, as follows

$$S(t) = \exp\{-A(t)\} = \exp\left\{-\int_0^t \alpha(s)ds\right\}. \tag{1.2}$$

1.1.2 Censoring and truncation

We suppose to be unable to observe the entire survival time T^*. For instance, the reason might be that the individual is still alive at the end of the study or the information about his status is lost during the study period. These are examples of right-censored times.

Let us denote by U the right-censoring time, which is the time from the origin to the end of the study or to the exit of the subject from the study for other reasons. When studying a group of subjects, time U may not be observed for each individual. This is the case when $T^* < U$. On the other hand, when the survival time is not observed for some individual, it means that $U < T^*$. Therefore, we define $T = \min(T^*, U)$ to be the follow-up time, which is an observable variable, and the indicator function $\Delta = I(T^* \leq U)$, which is equal to 1 if the survival time T^* is observed and equal to 0 if the observation is right-censored. The observation is then the pair (T, Δ). If we suppose to have a sample with n independent and identically distributed (i.i.d.) observations, the observed data are the pairs (T_i, Δ_i) for $i = 1, \ldots, n$, with $T_i = \min(T_i^*, U_i)$ and $\Delta_i = I(T_i^* \leq U_i)$.

There exist different types of right-censoring schemes (Andersen et al., 1993, Ander-

sen, 1998). The simplest one is called type I censorship and it happens when the study, and then the observation of subjects, ends at a common deterministic time u_e. Therefore, the right-censoring times are nonrandom and such that $U_i = u_e$ for $i = 1, \ldots, n$. Type II censorship is a scheme where the study ends at the time of the $r-$th failure, with $r \leq n$. In this case the right-censoring times are $U_i = T^*_{(r)}$ for $i = 1, \ldots, n$, where $T^*_{(i)}$ are the n ordered survival times, and the observed times T_i are dependent. The most common type of right-censoring scheme is the random censorship, where the censoring times U_i, $i = 1, \ldots, n$, are assumed to be i.i.d. with a given probability distribution. An important assumption in studies of survival analysis consists in the independent right-censoring, a situation where censoring times can be considered independent from the survival times (Kalbfleisch and Prentice, 2002). All the schemes previously described are independent right-censoring (see Section 1.3.2).

The basic question concerning survival data is how to incorporate incomplete observed data in order to obtain valid inference. Nonparametric and parametric inference is straightforward in case of independent right-censoring. More details are given later in the likelihood construction.

Another important kind of incomplete information is left-truncation (De Gruttola and Liao, 1998). In this case individuals enter in the study, and then they are observed, conditionally to not having experienced a certain event before the beginning of the study. Formally, survival times T^*_i are left-truncated if, given the times (random or nonrandom) V_i which represent entering into the study for each individual i, we observe $T^*_i | T^*_i > V_i$. Left-truncation is not to be confused with left-censoring. In the latter situation all individuals, both having and not having experienced a certain event before the beginning of the study, are observed, but for those experiencing the event before the beginning of the study, the only available information is that $T^*_i \leq V_i$.

1.1.3 The counting process notation

The general framework for survival data, which is given in the previous subsections, is common to all the literature about survival analysis. During the last two decades, the full development of the theory of martingales and counting processes has enabled most authors to work in the field of survival analysis using the counting process representation of the data (Fleming and Harrington, 1993, Andersen et al., 1993, Chap. 2).

Here we first present the counting process notation for complete data and then we gen-

eralize to the case of right-censored observations. Martingale theory will be presented later in Section 1.2.1.

Consider t as the time scale, varying in the time interval $(0, \tau]$. In literature, the end-point $\tau = \infty$ is often considered. The survival time T^* can be represented by the counting process

$$N(t) = I(T^* \leq t),$$

which assumes value zero until the jump to value one at time T^*. $N(t)$ is a stochastic process counting the number of observed events in the interval $(0, t]$. The martingale associated with the counting process $N(t)$ is defined as

$$M(t) = N(t) - \Lambda(t) \tag{1.3}$$

with compensator

$$\Lambda(t) = \int_0^t \lambda(s)\, ds. \tag{1.4}$$

M is generally called the counting process martingale. The term $\lambda(s)$ is the intensity process associated with the compensator, $\Lambda(t)$. This last term is also called the integrated or cumulative intensity process. The hazard in (1.1), which expresses a deterministic function, is linked to $\lambda(s)$ by

$$\lambda(t) = Y(t)\alpha(t).$$

The term $Y(t)$ is often called the at-risk process. It is a stochastic process defined as

$$Y(t) = I(t \leq T^*),$$

which indicates the at-risk state of an individual. It is equal to the unity while the individual is at risk, i.e., under observation before the event a time T^* has occurred, and zero afterwards. The difference between $\lambda(t)$ and $\alpha(t)$ is that the latter is a deterministic part, which is often modelled, while the former is a stochastic process, which expresses when the hazard rate is observed.

The above formulation includes easily right-censored data, because it allows the martingale theory to be still valid. In this case the counting process is

$$N(t) = I(T \leq t, \Delta = 1),$$

which is a right-continuous process with a jump to the unity only when the event is

observed at T^*, while the process is always equal to zero for a right-censored time. The at-risk process is $Y(t) = I(t \leq T)$, indicating that an individual is at risk until the event or the censorship occurs, and it is left-continuous and predictable.

For statistical purposes, in the counting process notation the n i.i.d. observations of a sample are the pairs of variables $(N_i(t), Y_i(t))$, for $i = 1, \ldots, n$, instead of the pairs (T_i, Δ_i) of the standard formulation. When analysing data, it is very common to have tied event times. As time is assumed to be continuous and it is desirable to assume an absolutely continuous distribution function for the survival times, it is convenient to handle the ties, breaking them according to different approaches (Efron, 1977, Therneau and Grambsch, 2000, Chap. 2), so that the problem is reformulated without ties. In fact, in case of no ties the theory of counting processes is kept in the easiest form, and inference is based on the simple process $N_i(t)$ with a possible jump of height 1.

A final remark concerns the usefulness of the counting process and martingale representation. The decomposition $N(t) = \Lambda(t) + M(t)$ can be thought of as the usual statistical form where the observed data are equal to the sum of the model and the error. The martingale process represents an error process, and $\Lambda(t) = \int_0^t Y(s)\alpha(s)\, ds$ expresses the expected number of events in $(0, t]$, which can be modelled by choosing a regression model for $\lambda(t)$, or, equivalently, a regression form for $\alpha(s)$. This comparison is also motivating the construction of residuals for goodness-of-fit methods based on martingales. A more formal justification to this interpretation is given by the asymptotic theory related to martingales, reviewed later in the present chapter.

1.2 Counting processes and martingale theory

The present section describes the basic concepts of counting processes and martingale theory in continuous time. The entire nonparametric and semiparametric theory for survival data and statistical modelling has relied on this theory during the last two decades. Moreover, the counting process representation allows us to generalize and extend the original basic regression models in survival analysis, handling the statistical formulation within a single comprehensive general framework. Some examples of useful extensions are time-varying coefficients and time-dependent covariates, analysis of residuals, multiple time scales, recurrent and multiple events.

We shall describe martingales and their properties, and formalize their connection to counting processes. Finally, we shall illustrate the theory in the specific context of

independent censoring (Andersen et al., 1993, Chap. 3).

1.2.1 Martingales

Our attention focuses on discrete events occurring in continuous time. Then, we consider time t within a given time interval $[0, \tau]$. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\mathcal{F}$ is a σ-field and P is a probability measure defined on $\mathcal{F}$. A stochastic process is defined as a family of random variables $\{Z(t), t \geq 0\}$ with sample paths $Z(t, \omega)$ for every $\omega \in \Omega$. The family of increasing sub-σ-fields, $\mathcal{F}_t = \sigma\{Z(s), 0 \leq s \leq t\}$, is called the filtration generated by the process Z.

A martingale with respect to a filtration $\mathcal{F}_t$ is an adapted right-continuous stochastic process M with left-hand limits (cadlag process) which is integrable, i.e.,

$$\mathrm{E}|M(t)| < \infty \qquad \text{for all } t,$$

and satisfies the martingale property

$$\mathrm{E}(M(t)\,|\mathcal{F}_s\,) = M(s) \quad \text{for all } s \leq t. \tag{1.5}$$

Property (1.5) states that information up to the present time s does not give further information about the expected value of M in the future time t. As the martingale property can be written equivalently as

$$\mathrm{E}(dM(t)\,|\mathcal{F}_{t-}\,) = 0 \quad \text{for all } t > 0, \tag{1.6}$$

where $dM(t) = M((t+dt)-) - M(t-)$, the martingale M has zero-mean increments given the past $\mathcal{F}_{t-}$.

Hereafter two important properties of a martingale are described:

- A martingale has constant mean in time, because it is $\mathrm{E}(M(t)) = \mathrm{E}(M(0))$, and if at the time origin it is $M(0) = 0$, then the mean of the martingale is zero, i.e., $\mathrm{E}(M(t)) = 0$ for all $t > 0$ (zero-mean martingale);

- the martingale increments are uncorrelated, i.e.,

$$\mathrm{Cov}(M(t) - M(t-s), M(t+u) - M(t)) = 0 \quad \text{for all } t, s, u \geq 0; \tag{1.7}$$

A process is a submartingale if in the property (1.5) the inequality holds, i.e.,

$$\mathrm{E}(M(t)\,|\mathcal{F}_s) \geq M(s) \quad \text{for all } s \leq t.$$

A martingale is square integrable when $\sup_{t\in[0,\tau]} \mathrm{E}(M(t)^2) < \infty$. If the martingale property holds locally for a process M, then M is called a local martingale.

In order to explain further properties and theorems related to martingales, we need to define formally a predictable process H and a compensator. A stochastic process H is called predictable if it is measurable with respect to the $\sigma-$algebra generated by the adapted processes whose paths are left-continuous. In this sense, all the left-continuous adapted processes are predictable, and also any deterministic measurable function. Given a cadlag adapted process X, a compensator $\tilde{X}$ is a predictable, cadlag and finite variation process such that $X - \tilde{X}$ is a local zero-mean martingale. If a compensator exists, it is unique.

Our intent is to be able to write a stochastic process as the sum of a martingale and a predictable process, in order to justify the decomposition of the counting process mentioned in (1.3). The definitions previously introduced in the current subsection are now needed to explain for which processes the decomposition holds. More formally, the answer is contained in a crucial theorem which states the so-called Doob-Meyer decomposition.

Theorem 1.2.1 *Let X be a cadlag adapted process. Then X has a compensator if and only if X is the difference of two local submartingales.*

For further details, see Andersen et al. (1993, Chap. 2).

As a consequence, if X is also a local submartingale, then it has a compensator, since X is the difference of two local submartingales, X itself and the trivial constant process 0.

Suppose M is a local square integrable martingale. Then, by Jensen's inequality, M^2 is a local submartingale and therefore, by the just mentioned consequence of theorem 1.2.1, it has a compensator. The compensator of M^2 is called the predictable variation process of M and it is denoted by $\langle M\rangle$. Consequently, we have that $M^2 - \langle M\rangle$ is a local zero-mean martingale. The predictable variation process of M is the limit in probability of the approximations

$$\sum_j \mathrm{E}\left[(M(t_{j+1}) - M(t_j))^2\,|\mathcal{F}_{t_j}\right] = \sum_j \mathrm{Var}\left[M(t_{j+1}) - M(t_j)|\mathcal{F}_{t_j}\right] \qquad (1.8)$$

for increasingly fine partitions of the interval $[0,t]$, $0 = t_0 < t_1 < \cdots < t_j < t_{j+1} < \cdots < t_n = t$.

A similar explanation is related to the so-called predictable covariation process $\langle M, \tilde{M} \rangle$ of M and $\tilde{M}$, which is the compensator of the product $M\tilde{M}$, with M and $\tilde{M}$ being two local square integrable martingales. In general, the predictable covariation process is useful in the asymptotic theory to identify asymptotic covariances in the statistical problems, since

$$\text{Cov}(M(s), \tilde{M}(t)) = \text{E}(\langle M, \tilde{M} \rangle)(t), \quad s \leq t. \tag{1.9}$$

Another important process to consider in this context is the optional variation process of M, denoted by $[M]$. It is the limit in probability of the sums of squares

$$\sum_j \left(M(t_{j+1}) - M(t_j)\right)^2, \tag{1.10}$$

for increasingly fine partitions of the interval $[0,t]$, $0 = t_0 < t_1 < \cdots < t_j < t_{j+1} < \cdots < t_n = t$. This process is defined for M being just a local martingale, and not anymore a local square integrable martingale. When M has finite variation, the optional variation process has the explicit form $[M](t) = \sum_{s \leq t} \left[M(s) - M(s-)\right]^2$. The process $M^2 - [M]$ is a local martingale.

The process $[M]$, unlike $\langle M \rangle$, is not predictable. Moreover, in statistical applications $\langle M \rangle$ is determined by the model characteristics, as suggested by the approximations in (1.8). The process $[M]$ may instead be computed from the data, as seen from (1.10). Finally, if $[M]$ is locally integrable, $\langle M \rangle$ is the compensator of $[M]$. Therefore, we are able to compute both the predictable and optional variation processes for a statistical application on the basis of (1.10).

Our attention concentrates on statistical problems where stochastic integrals related to a martingale M can be computed easily. The reason is that these stochastic integrals have a pathwise interpretation, more specifically, they are ordinary pathwise Lebesgue integrals. A special property arises when the integrand is a predictable process H and we integrate with respect to a local martingale M. The resulting process $\int H dM$ is a local martingale, and its predictable and optional variation processes can be obtained from $\langle M \rangle$ and $[M]$ by the formulas

$$\Big\langle \int H\, dM \Big\rangle = \int H^2\, d\langle M \rangle \quad \Big[\int H\, dM \Big] = \int H^2\, d[M]. \tag{1.11}$$

For further details and results on martingale theory related to the statistical analysis of survival data, see Fleming and Harrington (1993).

1.2.2 Counting processes

In the present subsection a counting process will be formally defined. The principal results and properties will be described by applying the martingale theory from the previous subsection to counting processes.

A stochastic process $N(t)$ is called a counting process if it is adapted to the filtration $\mathcal{F}_t$, cadlag, almost surely finite for all t, with $N(0) = 0$ and with piecewise constant paths having jumps of size 1.

Because of its definition, a counting process $N(t)$ is a local submartingale. Therefore, as explained in the previous subsection, it has a compensator, called Λ. The process $\Lambda(t)$ is nondecreasing, predictable and with $\Lambda(0) = 0$. Moreover, because of the definition of a compensator, the process $M = N - \Lambda$ is a local zeromean martingale with respect to $\mathcal{F}_t$. Furthermore, if $\mathrm{E}(\Lambda(t) < \infty$, then M is a martingale, as it verifies all the martingale conditions.

An important property of the counting process is that

$$\mathrm{E}(N(t)) = \mathrm{E}(\Lambda(t)) \tag{1.12}$$

as M is a zeromean martingale. A martingale increment is defined as $dM(t) = M((t + dt)-) - M(t-)$, and the increments dN and $d\Lambda$ are defined similarly. Thus, the just mentioned property, as in general all the other properties about martingales and counting processes, can be written in the form $\mathrm{E}(dN(t)|\mathcal{F}_{t-}) = E(d\Lambda(t)|\mathcal{F}_{t-})$. This is immediately obtained from the decomposition of the martingale increments, $dM(t) = dN(t) - \lambda(t)dt$, and the fact that $dM(t)$ has zeromean.

In case we restrict our attention to the case of independent censoring and absolutely continuous distributions for the survival times, a fundamental consequence follows. First, we stress that the continuity gives the relation between the cumulative intensity process Λ and the intensity process λ, expressed in (1.4). Therefore, equation (1.12) becomes

$$\mathrm{E}(dN(t)|\mathcal{F}_{t-}) = \lambda(t)dt. \tag{1.13}$$

This conclusion is obtained because the predictable process $\lambda(t)dt$ is a nonrandom term, given $\mathcal{F}_{t-}$.

Hereafter, we mention two important statements concerning the variance of a counting processes martingale:

$$\text{Var}(dM(t)|\mathcal{F}_{t-}) = d\langle M\rangle(t), \tag{1.14}$$

since $\text{Var}(dM(t)|\mathcal{F}_{t-}) = \text{E}[(dM(t))^2|\mathcal{F}_{t-}] = \text{E}[d(M^2)(t)|\mathcal{F}_{t-}]$ and $d\langle M\rangle$ is the compensator of dM^2;

$$\text{Var}(dM(t)|\mathcal{F}_{t-}) = d\Lambda(t)[1 - d\Lambda(t)] \approx d\Lambda(t). \tag{1.15}$$

The explanation of (1.15) consists of two facts. First, $dM(t) = dN(t) - \text{E}(dN(t)|\mathcal{F}_{t-})$, i.e. the martingale increment is the difference between the counting process increment and its conditional expectation. Second, N is a process assuming only two possible values, 0 or 1, and its definition states that N has jumps of size 1. Hence, $\text{Var}(dM(t)) = \text{Var}(dN(t))$.

The expressions in (1.14) and (1.15), if observed together, lead heuristically to assess that

$$\langle M\rangle(t) = \Lambda(t), \tag{1.16}$$

i.e., the predictable variation process of M is just the compensator Λ of the counting process. The following result for the optional variation process can be formally obtained from the martingale theory previously described (Section 1.2.1):

$$[M](t) = N(t).$$

Moreover, equation (1.16) arises formally from noting that $[M]$ is locally integrable and hence $\langle M\rangle$ is its compensator.

In statistical problems we are faced with multivariate counting processes, as inference is based on a sample of size n. A multivariate counting processes,

$$\boldsymbol{N} = (N_1, \ldots, N_j, \ldots, N_n), \tag{1.17}$$

is a vector of counting processes, each of them defined as previously in the current subsection, and such that they do not have simultaneous jumps. Each process N_j is associated with the counting process martingale M_j and the compensator Λ_j, as in (1.3), and it has all the properties and results previously described for the one-dimensional counting processes. The additional property concerns the orthogonality of martingales, that is the predictable covariation process is null for each pair M_j, M_l, with $j \neq l$. This fact leads to the compacting matrix notation $\langle \boldsymbol{M}\rangle = \Lambda$ and $[\boldsymbol{M}] =$

$\boldsymbol{N}$.

In case of counting processes, specific expressions about stochastic integration arise from equations (1.11) (Andersen et al., 1993, Chap. 2).

1.2.3 Asymptotic theory

Most of the asymptotic properties of estimators based on martingales in counting process models arise from a central limit theorem for martingales. There exist many versions of the theorem, which anyway generalize the original version given by Rebolledo (1980). Before illustrating the theorem, we briefly introduce some necessary concepts.

Let us consider a vector of k $\mathbb{R}$-valued local square integrable martingales, denoted by $M^{(n)} = (M_1^{(n)}, \ldots, M_k^{(n)})$, where n represents the sample size. Let $\{M^{(n)}(t) : t \in [0, \tau]\}$ be a sequence for $n = 1, 2, \ldots$. For each $M_h^{(n)}$ in $M^{(n)}$, with $h = 1, \ldots, k$, let $M_{\epsilon h}^{(n)}$ be the corresponding martingale containing all the jumps of $M_h^{(n)}$ larger in absolute value than ϵ. That is, all the jumps of $M_h^{(n)}$ are such that $|M_{\epsilon h}^{(n)}(s) - M_{\epsilon h}^{(n)}(s-)| > \epsilon$ for $s \leq t$.

An $\mathbb{R}^k$-valued martingale U is said to be Gaussian if it has continuous sample paths, $U(0) = 0$, and any finite family $(U(t_1), \ldots, U(t_j))$ has Gaussian distribution. The covariance matrix of $U(t)$, $V(t)$, is such that the increment $V(t) - V(s)$, for $s \leq t$, is positive semidefinite. Moreover, the Gaussian martingale increment $U(t) - U(s)$ has a normal distribution $N(0, V(t) - V(s))$ and is independent of $(U(l); l \leq s)$ for $s \leq t$.

Theorem 1.2.2 *Central limit theorem for martingales.*
If $(M^{(n)}(t) : t \in [0, \tau])$ is a sequence of $\mathbb{R}^k$-valued local square integrable martingales and the following conditions

$$\langle M^{(n)} \rangle(t) \xrightarrow{P} V(t) \qquad \textit{for all } t \textit{ as } n \to \infty, \tag{1.18}$$

$$\langle M_{\epsilon l}^{(n)} \rangle(t) \xrightarrow{P} 0 \qquad \textit{for all } t, l \textit{ and } \epsilon > 0 \textit{ as } n \to \infty \tag{1.19}$$

hold, then

$$M^{(n)} \xrightarrow{\mathcal{D}} U \textit{ as } n \to \infty, \tag{1.20}$$

i.e., the process $M^{(n)}$ converges weakly to a Gaussian martingale U with covariance function V. Moreover, $\langle M^{(n)} \rangle$ and $[M^{(n)}]$ converge uniformly on compact subsets of

$[0, \tau]$, *in probability, to* V.

The weak convergence refers to the space $\{\mathcal{D}([0,\tau])\}$ of the $\mathbb{R}^k$-valued cadlag functions on $[0,\tau]$, with the Skorohod topology as defined in Billingsley (1968). The theorem assesses that the jumps of $M^{(n)}$ become negligible as $n \to \infty$, and thus the process has asymptotically continuous sample paths, and the predictable variation process, which is equal to the compensator in case of counting processes, converges in probability to a deterministic function.

By the application of Theorem 1.2.2 to the martingales or to functionals of martingales in the counting process setting, it is possible to determine the asymptotic distributions of many estimators and use these results for defining tests of hypotheses and confidence intervals.

1.3 Model specification for counting processes

In this section the model specification is presented within the counting process setting, first in the case of complete data, and then for right-censored survival times. In order to illustrate how the likelihood function is constructed, we describe the simplest case of a single uncensored survival time and then we generalize to a random sample of survival data. Moreover, it is shown how to accommodate the likelihood function to incomplete information due to right-censorship. When a regression model for the hazard function is desired, the likelihood is then a function of the regression parameters in the distribution of survival times, besides the parameters in the distribution of right-censored times.

1.3.1 Likelihood and partial likelihood construction

As a first step, the simple case of a single complete observation is considered. Let T^* be the survival time in $[0,\tau]$ with density function f^θ depending on a parameter θ, which might have finite or infinite dimension. Thus, we denote by $\alpha^\theta(t)$ the corresponding hazard function. The counting process N associated with T^* is univariate and the compensator $\Lambda^\theta(t) = \int_0^t \lambda^\theta(s)ds = \int_0^t Y(s)\alpha^\theta(s)ds$ represents the cumulative intensity process, with $Y(t) = I(T^* \geq t)$ being the at-risk process.

The likelihood function for θ up to time t is given by

$$L(\theta, t) = \prod_{s \leq t, s \in [0,\tau]} (\lambda^\theta(s))^{\Delta N(s)} \exp\left(- \int_0^t \lambda^\theta(u) du \right), \tag{1.21}$$

where $\Delta N(t) = N(t) - N(t-)$.

The likelihood computed up to the entire interval $[0, \tau]$ is

$$\begin{aligned} L(\theta) &= \prod_t \left\{ Y(u)\alpha^\theta(u) \right\}^{\Delta N(u)} \exp\left(- \int_0^\tau Y(u)\alpha^\theta(u) du \right) \\ &= \alpha^\theta(T^*) \exp\left(- \int_0^{T^*} \alpha^\theta(u) du \right), \end{aligned} \tag{1.22}$$

which reduces to the density function at T^*, since $L(\theta) = \alpha^\theta(T^*) S^\theta(T^*) = f^\theta(T^*)$.

We now generalize the previous case to a multivariate counting process for uncensored survival data. Let $T_1^*, \ldots, T_n^*$ be independent survival times with hazard functions $\alpha_i^\theta, i = 1, \ldots, n$. $\boldsymbol{N} = (N_1, \ldots, N_n)$ is the associated multivariate counting process with intensity process $\boldsymbol{\lambda^\theta} = (\lambda_1^\theta, \ldots, \lambda_n^\theta)$. The likelihood function for θ is then given by

$$L(\theta) = \prod_{i=1}^n \prod_t (\lambda_i^\theta(t))^{\Delta N_i(t)} \exp\left(- \sum_{i=1}^n \int_0^\tau \lambda_i^\theta(u) du \right). \tag{1.23}$$

Similarly to above the likelihood can be written as

$$L(\theta) = \prod_{i=1}^n \alpha_i^\theta(T_i^*) \exp\left(- \sum_{i=1}^n \int_0^{T_i^*} \alpha_i^\theta(u) du \right), \tag{1.24}$$

since, recalling the independence of survival times, it verifies

$$L(\theta) = \prod_{i=1}^n \alpha_i^\theta(T_i^*) S_i^\theta(T_i^*) = \prod_{i=1}^n f_i^\theta(T_i^*).$$

If the statistical model is specified by regression on covariates, we need to observe the (T_i^*, X_i), for $i = 1, \ldots, n$. The X_i is the covariate vector for individual i. Assume that $T_1^*, \ldots, T_n^*$ are independent conditionally on the covariates in $\boldsymbol{X} = (X_1, \ldots, X_n)$, and that the conditional distribution of $T_i^* | X_i$ has hazard $\alpha_i^\theta(t)$ which depends on a parameter θ. Moreover, suppose that the marginal distribution of $\boldsymbol{X}$, $P_{\phi\theta}$ depend on a nuisance parameter ϕ and on θ. Therefore, each compensator $\Lambda_i(\cdot)$ associated with the counting process $N_i(\cdot)$ depends only on the parameter of interest θ. The likelihood

function for the parameters (θ, ϕ) can be factorized in the form

$$L(\theta, \phi) = L_0(\theta, \phi) L_\tau(\theta), \tag{1.25}$$

where $L_0(\theta, \phi) = P_{\phi\theta}(\boldsymbol{X})$. The remaining factor $L_\tau(\theta)$ is the partial likelihood for θ and is given by (1.23), or equivalently by (1.24). It represents the conditional distribution of the $T_i^* | \boldsymbol{X}$, $i = 1, \ldots, n$, evaluated at $T_1^*, \ldots, T_n^*$. If the distribution of $\boldsymbol{X}$ depends only on ϕ, that is $L_0(\theta, \phi) = L_0(\phi)$, then $L_\tau(\theta)$ is a full likelihood for θ for each given ϕ in its parameter space.

In the literature there exist various regression models for the hazard function, synthetically written as $\alpha_i^\theta(t) = g(t, \theta, x_i)$ with x_i being the observed values of x_i. Essentially they differ with respect to the link function $g(\cdot)$ between hazard and covariates. The most common examples are the multiplicative and additive hazards models, where the link functions $g(\cdot)$ are, respectively, in multiplicative or additive form. For these models, inference is simply based on the (partial) likelihood $L_\tau(\theta)$ where $\alpha_i^\theta(t)$ is substituted by its regression form $g(t, \theta, x_i)$.

1.3.2 Right-censorship

This section illustrates briefly how the counting process modifies in the presence of right-censorship and it serves as an introduction to the following section about model specification for right-censored data.

Given $(T_1^*, \ldots, T_n^*)$ independent survival times, we consider a multivariate process $\boldsymbol{N}^* = (N_1^*, \ldots, N_n^*)$ adapted to the filtration $\mathcal{F}_t^*$ on the probability space $(\Omega, \mathcal{F}^*, P)$. The simplest situation is when the filtration is the natural one generated by the counting process itself, i.e., $\mathcal{F}_t^* = \mathcal{N}_t = \sigma(\boldsymbol{N}(s) : s \leq t)$, while in regression models, the filtration $\mathcal{F}_t^*$ may also incorporate information about covariates. When we are faced with incomplete (right-censored) survival data, the counting process $\boldsymbol{N}^*$ can not be completely observed. Therefore, partially observed counting processes are defined as

$$N_i(t) = \int_0^t C_i(u) dN_i^*(u), \quad i = 1, \ldots, n, \tag{1.26}$$

and called right-censored counting processes, where

$$C_i(t) = I(t \leq U_i)$$

is the so-called individual right-censoring process (U_i is the time of censorship for

individual i).

The right-censoring process $C_1(t), \ldots, C_n(t)$ are not adapted to the filtration $\mathcal{F}_t^*$ and for this reason we consider the enlarged filtration

$$\mathcal{G}_t^* = \mathcal{F}_t^* \vee \sigma(C_i(u); i = 1, \ldots, n; u \leq t),$$

so that $C_1(t), \ldots, C_n(t)$ are adapted and predictable with respect to $\mathcal{G}_t^*$. Moreover, for any individual i either N_i^* or C_i is not fully observed. Hence, the filtration $\mathcal{G}_t^*$ can not be fully observed and therefore, we need to work with a reduced filtration $\mathcal{F}_t$, which is generated by the observed data.

Given $\boldsymbol{N}$, the corresponding intensity process $\boldsymbol{\lambda}$ with respect to $\mathcal{F}_t$ might be different from the one associated with the entire counting process $\boldsymbol{N}^*$, $\boldsymbol{\lambda}^*$, since it may be changed by the right-censoring mechanism. When this does not occur, i.e., $\boldsymbol{\lambda}^*(t) = \boldsymbol{\lambda}(t)$, then we have independent right-censoring. A formal definition of independent right-censoring is given by Andersen et al. (1993, Chap. 3): The right-censoring mechanism leading to the observable counting process $\boldsymbol{N}$ generated by the C_i is said to be independent if the compensator of $\boldsymbol{N}^*$ with respect to $\mathcal{G}_t^*$ is equivalent to the compensator λ^* with respect to $\mathcal{F}_t^*$.

The intensity processes for subjects being at risk at any time t are unchanged by the modified filtration $\mathcal{G}_t^*$ due to right-censoring. However, our interest is on the intensity process $\boldsymbol{\lambda}$ of the observed $\boldsymbol{N}$ with respect to $\mathcal{F}_t$. If $C_i(t)\lambda_i^*(t)$, for $i = 1, \ldots, n$, are predictable with respect to $\mathcal{F}_t$, then it follows from the previous definition of right-censoring that the intensity processes of $\boldsymbol{N}$ with respect to $\mathcal{F}_t$ are

$$\lambda_i(t) = C_i(t)\lambda_i^*(t). \tag{1.27}$$

In other words, when the right-censoring is independent, the information carried by the right-censoring process $C_i(t)$ does not modify the intensity process for $\boldsymbol{N}$ at time t. The interpretation is that the risk set, containing individuals being at risk just before a certain time t, is representative of what the sample of individuals would have been without censoring.

Let $Y_i^* = I(t \leq T_i^*)$ be the risk indicator for $\boldsymbol{N}^*$. Independent right-censoring modifies the risk indicator so that

$$Y_i(t) = C_i(t)Y_i^*(t) = I(t \leq T_i), \tag{1.28}$$

with $T_i = \min(T_i^*, U_i)$ being the observed right-censored survival time for individual i. Therefore for the intensity process, we have $\lambda_i(t) = Y_i(t)\alpha_i(t)$, with $\alpha_i(t)$ being the hazard function for individual i in an independent right-censoring scheme.

1.3.3 Model specification for right-censored data

In this section we describe the likelihood function under the assumption of independent right-censoring and we keep the same notation used in Section 1.3.2.

Consider a right-censored counting process $\boldsymbol{N}$, with functions $N_i(t)$ as in (1.26), the at-risk indicators $Y_i(t)$ for $i = 1, \ldots, n$, defined as in (1.28). Suppose a model for the hazard function $\alpha^\theta(t)$ is specified depending on a parameter θ. Therefore, the intensity process associated with the right-censored counting process $N_i(t)$ is given in (1.27) and it depends also on θ, since $\lambda_i^\theta(t) = Y_i(t)\alpha_i^\theta(t)$.

The likelihood function can be constructed starting from the factorization $L(\theta, \phi) = L_\tau^u(\theta, \phi)L_\tau^c(\theta)$, similarly to what was done in (1.25) for uncensored observations. The first factor $L_\tau^u(\theta, \phi)$ may contain information about the additional parameter ϕ related to the distribution of the censoring variables U_i or/and the distribution of a possible covariate vector $\boldsymbol{X}$. The second factor is the likelihood for θ and has the form

$$\begin{aligned} L_\tau^c(\theta) &= \prod_{i=1}^n \left\{ \prod_t (\lambda_i^\theta(t))^{\Delta N_i(t)} \exp\left(- \int_0^\tau \lambda_i^\theta(u)du \right) \right\} \\ &= \prod_{i=1}^n \alpha_i^\theta(T_i) \exp\left(- \sum_{i=1}^n \int_0^{T_i} \alpha_i^\theta(u)du \right), \end{aligned} \tag{1.29}$$

with $T_i = \min(T_i^*, U_i)$ being the observed time for individual i. The expression of $L_\tau^c(\theta)$ in (1.29) represents a partial likelihood for θ, and contains the terms related to the distribution of the right-censored survival times T_i, $i = 1, \ldots, n$.

Presence of independent right-censoring does not alter the form of the partial likelihood for θ, as it can be noted comparing expressions in (1.29) and (1.24). If the first factor $L_\tau^u(\theta, \phi)$ does not depend on the parameter of interest θ, then $L_\tau^c(\theta)$ corresponds to a full likelihood for θ at each fixed ϕ in its parameter space and the right-censoring mechanism is said to be noninformative for θ.

1.3.4 Maximum partial likelihood estimation

The basic inference based on the maximum partial likelihood estimator is presented in this section. The theory is valid for i.i.d. observations of counting processes $N_i(t)$, $i = 1, \ldots, n$, under independent right-censorship, but can also be used in a more general context. In this section we assume that θ is a p-dimensional parameter.

The log-partial likelihood function for θ can be written as

$$l_\tau^c(\theta) = \sum_i \left\{ \int_0^\tau \log(\lambda_i^\theta(t)) dN_i(t) - \int_0^\tau \lambda_i^\theta(t) dt \right\},$$

which yields the p-dimensional score function

$$U_\tau(\theta) = \sum_i \left\{ \int_0^\tau \frac{\partial}{\partial\theta} \log(\lambda_i^\theta(t)) dN_i(t) - \int_0^\tau \frac{\partial}{\partial\theta} \lambda_i^\theta(t) dt \right\}.$$

The maximum likelihood estimator for θ is given as a solution to the equation $U_\tau(\theta) = 0$. It can be proved that there exists a consistent estimator $\hat{\theta}$, and that $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with covariance matrix $\mathcal{I}^{-1}(\theta_0)$, where θ_0 is the true parameter value. The information matrix $\mathcal{I}$ has elements

$$\mathcal{I}_{j,l}(\theta_0) = \mathrm{E}\big(- \frac{\partial^2}{\partial\theta_j \partial\theta_l} l_\tau^c(\theta)\big), \qquad j, l = 1, \ldots, p,$$

evaluated at θ_0. The asymptotic covariance matrix may be estimated by the observed information at $\hat{\theta}$, $\hat{\mathcal{I}}(\hat{\theta})$, with elements

$$\hat{\mathcal{I}}_{j,l}(\theta_0) = n^{-1} \sum_i \left\{ - \int_0^\tau \frac{\partial^2}{\partial\theta_j \, \partial\theta_l} \log(\lambda_i^*(s)) dN_i(s) + \int_0^\tau \frac{\partial^2}{\partial\theta_j \, \partial\theta_l} \lambda_i^*(s) ds \right\}.$$

1.3.5 Regression models for incomplete survival data

We describe briefly some essential regression models using the counting process representation. Let $\lambda^\theta(t)$ be the intensity process associated with the counting process $N(t)$, with $t \in [0, \tau]$, and θ the parameter of interest. Let $X(t)$ be a vector of covariates, possibly time-dependent. A regression model for the intensity process $\lambda^\theta(t)$ may be specified by choosing a functional form for the hazard function $\alpha^\theta(t)$, since $\lambda^\theta(t) = Y(t)\alpha^\theta(t)$ and the at-risk indicator $Y(t)$ does not depend on any parameter. The two most common classes of regression models are based on an additive or multiplicative form, and are called additive hazards models and multiplicative hazards

models, respectively.

Example 1.3.1 An additive hazards model is a nonparametric model specified by the following form for the intensity process (Aalen, 1980)

$$\lambda^{\beta}(t) = Y(t)X^T(t)\beta(t), \tag{1.30}$$

where $X(t) = (X_1(t), ..., X_p(t))$ is a p-dimensional vector of covariates. In this model, known as the additive Aalen model, the parameter of interest θ is simply the locally integrable p-dimensional parameter $\beta = (\beta_1(t), ..., \beta_p(t))$.

Inference for this model is usually made on the cumulative regression coefficient $B(t) = \int_0^t \beta(s)ds$, and it is based on the counting process martingale decomposition

$$N(t) = \int_0^t X(s)\beta(s)ds + M(t).$$

Estimators for the cumulative regression coefficient $B(t)$ can be obtained either by least square methods for multiple linear regression (Aalen, 1980) or by maximum partial likelihood methods (Sasieni, 1992, Greenwood and Wefelmeyer, 1991).

The multiplicative hazards models are semiparametric models, where the effects of covariates on the hazard function follow a multiplicative scale. The most famous example of this class is the proportional hazards model, known also as the Cox model.

Example 1.3.2 In the Cox model (Cox, 1972), the intensity process is specified as follows

$$\lambda^{\theta}(t) = Y(t)\alpha_0(t)\exp(X^T(t)\beta), \tag{1.31}$$

where $\theta = (\alpha_0, \beta)$ with α_0 being a nonparametric locally integrable function and $\beta = (\beta_1, \ldots, \beta_p)$ being the p-dimensional vector of regression coefficients. $X(t)$ is here the p-dimensional vector $(X_1(t), ..., X_p(t))$ of covariates. The parameter $\lambda_0(t)$ is denoted as the baseline hazard function.

Given i.i.d. observations $(N_i(t), Y_i(t), X_i(t))$, $i = 1, \ldots, n$, in the time interval $[0, \tau]$, inference for the Cox model is based on the well-known partial likelihood function (Cox, 1972) for β

$$L(\beta) = \prod_t \prod_i \left(\frac{\exp(X_i^T(t)\beta)}{S_0(t, \beta)} \right)^{\Delta N_i(t)}, \tag{1.32}$$

with

$$S_0(t,\beta) = \sum_i Y_i(t)\exp(X_i^T(t)\beta).$$

The expression in (1.32) is found by replacing the so-called cumulative baseline hazard $A_0(t) = \int_0^t \alpha_0(u)du$ with its estimator in the likelihood function for θ (1.29), where the intensity process $\lambda^\theta(t)$ has the regression form (1.31).

Another example of multiplicative hazards models is an extension of (1.31) in Example 1.3.2, where the regression coefficients are allowed to vary over time so that are of the form $\beta(t)$. For an introduction to this model, see Martinussen and Scheike (2006, Chap. 6) and references therein.

Recently, regression models that combines the multiplicative and additive intensity models, have been proposed in the literature, leading to a more general setting where both the Cox model and the additive Aalen model are included as special cases. This combination has been studied following various approaches, the most relevant being the ones by Lin and Ying (1995), Martinussen and Scheike (2002) and Scheike and Zhang (2002).

1.4 Competing risks

In this section a general overview of multi-state models is presented as a framework for the competing risks setting. In the literature, there exists an alternative approach to competing risks, denoted as latent failure times approach, which assumes a certain number of potential failure times for each individuals. There is a vast literature on this latter approach, especially in applied areas other than biostatistics. A complete picture of the latent failure times approach for competing risks is given by Tsiatis (1998) and a brief discussion about problems of nonidentifiability of the survival distribution related to this approach is found in Tsiatis (1975).

1.4.1 Multi-state models

A multi-state model is a model for a stochastic process in continuous time with a finite number of states. Generally, a time origin is given and set equal to 0. Individuals under study are observed in time and they can experience one or multiple events, each one corresponding to a state of the process. The states might represent different aspects of the history of individuals, according to the problem studied. For example,

in biomedicine, they can be different causes of death from a certain disease, distinct phases of the disease, clinical symptoms or marginal side-effects.

There is a vast literature in this field, and various multi-state models are presented in applied contexts (Courgeau and Lelièvre, 1992, Commenges, 1999). The general theory is presented in Andersen and Keiding (2002), Andersen et al. (1993) and Hougaard (1999, 2000). Various modelling approaches within the multi-state setting, related inference and softwares are reviewed in Meira-Machado et al. (2007).

Hereafter a formal description of a multi-state model is presented. This model consists of a stochastic process, denoted here by $\{Z(t), t \in [0, \tau]\}$, with right-continuous paths and finite state space $\mathcal{S} = \{0, 1, \ldots, k\}$. Let $(\Omega, \mathcal{F}, P)$ be the reference probability space, where $\mathcal{F}$ is the filtration generated by the process $Z(t)$.

The distribution of the process is determined by the matrix $P(s,t)$, for $s, t \in [0, \tau]$, of transition probabilities between states

$$P_{hl}(s,t) = P(Z(t) = l | Z(s) = h, \mathcal{F}_{s-}) \qquad h, l \in \mathcal{S} \quad s \leq t, \tag{1.33}$$

or, equivalently, by the matrix $Q(t)$ of transition intensities

$$\alpha_{hl}(t) = \lim_{\Delta t \to 0} \frac{P_{hl}(t, t + \Delta t)}{\Delta t} \qquad h, l \in \mathcal{S}, \quad h \neq l, \quad t \in [0, \tau), \tag{1.34}$$

which are supposed to exist. The α_{hh}, for $h \in \mathcal{S}$, are defined as $\alpha_{hh} = -\sum_{h \neq l} \alpha_{hl}$, since the sum of the probabilities P_{hl} over $l = 0, 1, \ldots, k$ must be equal to one.

The initial distribution of the process is denoted by $\pi_h(0)$, for every $h \in \mathcal{S}$, and represents the probability to be in the state h at the time origin $t = 0$. The state probabilities are defined by the sum of the transition probabilities over the origin states, weighted by the initial distribution, i.e., $\alpha_l(t) = \sum_{h \in \mathcal{S}} P_{hl}(0,t)\pi_h(0)$, for $l \in \mathcal{S}$ and $t \in [0, \tau]$. A state h is absorbing when $\alpha_{hl, h \neq l}(t) = 0$ for all $t \in [0, \tau]$ and $l \in \mathcal{S}$, otherwise it is a transient state.

The multi-state model is built by associating statistical models to the transition intensities $\alpha_{hl}(t)$ in (1.34). Most models treated in the literature are associated with the intensity transitions of a nonhomogeneous Markov process. The homogeneity assumption is present if the transition intensities are constant in time and therefore the matrix Q is independent of time. Otherwise, that is when $\alpha_{hl}(t)$ depends on t, the process is nonhomogeneous. The Markov property states that

$$P_{hl}(s,t) = P(Z(t) = l | Z(s) = h, \mathcal{F}_{s-}) = P(Z(t) = l | Z(s) = h),$$

i.e., $\alpha_{hl}(t)$ depends on the history $\mathcal{F}_t$ only through the current state at time t.

Example 1.4.1 The simplest multi-state model is associated with a process with a single transient state 0 and a single absorbing state 1. In this situation, the survival time T of an individual represents the time from the origin (state 0) to the occurrence of a certain event (state 1). The survival probability is then given by the transition probability of still being in state 0 at time t, i.e., $S(t) = P_{00}(0,t) = \exp\left\{-\int_0^t \alpha(s)ds\right\}$, with a single transition intensity $\alpha_{01}(\cdot) = \alpha(\cdot)$.

The transition intensity $\alpha(\cdot)$ is the hazard function defined in (1.1) at the beginning of the chapter. Hence, a statistical model can be chosen for $\alpha(\cdot)$, under some general assumption. The Cox model (Example 1.3.2) is the simplest example of a regression model for the intensity transition $\alpha(\cdot)$, under the assumption of a nonhomogeneous Markov process.

1.4.2 Nonhomogeneous Markov multi-state models

For a Markov model, explicit expressions for the transition probabilities can be found by solving forward Kolmogorov differential equations (Sidney, 1992). They are functions of the transition intensities, and therefore of the hazard functions of the model.

We introduce here the integrated intensities $A_{hl}(t) = \int_0^t \alpha_{hl}(s)ds$, which are simply the cumulative hazards functions defined in Subsection 1.1.1. We use the notation dA_{hl} instead of α_{hl} to denote the entries of Q, referring to a more general context than absolutely continuous A_{hl}.

An important instrument for statistical inference is the so-called product integral representation for the transition probabilities

$$P(s,t) = \prod_{u\in(s,t]}(I + Q(u)), \tag{1.35}$$

where I is the identity matrix, and the product integral is defined by

$$\prod_{u\in(s,t]}(I + Q(u)) = \lim_{\max|t_v - t_{v-1}|\to 0}\prod(I + Q(t_v) - Q(t_{v-1})),$$

where $s = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[s,t]$. The theory about this representation is thoroughly explained in Andersen et al. (1993).

To understand the idea behind the product integral representation, let us consider the case when the A_{hl} are step functions. The corresponding process is then a Markov

chain in discrete time. The transition matrix $P(s,t)$ can be written as the product of transition matrices at each jump time between s and t, since the limit is obtained at the finite partition constructed from the jump times. The representation (1.35) can be considered as a generalization of this example to an infinitesimal fine partition.

When a Markov regression multi-state model is suitable, particular care needs to be taken when many states are present. In this last situation, all the transitions between states need to be modelled by regression of the corresponding transition intensities. The consequence is that a large number of regression coefficients must be estimated, and that there is the demand for a sufficiently large size sample of individuals. This problem requires less attention for a regression competing risks model, which will be described in the next section.

1.4.3 Counting process notation for multi-state models

Before introducing the competing risks model as a particular multi-state model, we briefly mention the counting process notation. We assume a nonhomogeneuous multi-state model based on a Markov process $Z(t)$. We consider n independent observations from this process over the time interval $[0,\tau]$ and denote them by $Z_i(t)$, for $i = 1,\ldots,n$. Moreover, we assume here independent right censoring or left truncation.

Hence, the counting process representation leads to defining the multivariate counting process $N = \{N_{hl}, h \neq l\}$, as in (1.17). Here

$$N_{hl} = \sum_{i=1}^{n} N_{i,hl}(t), \tag{1.36}$$

where $N_{i,hl}(t)$ is the process associated with individual i, which counts the number of observed direct transitions of the process $Z_i(t)$ from the state h to the state l. N has intensity process $\lambda = \{\lambda_{hl}, h \neq l\}$, where each element has the form $\lambda_{hl}(t) = Y_h(t)\alpha_{hl}(t)$. The at-risk process $Y_h(t)$ represents the number of individuals at risk in state h at time $t-$, and it is given by

$$Y_h(t) = \sum_{i=1}^{n} Y_{i,h}(t), \tag{1.37}$$

with $Y_{i,h}(t) = I(Z_i(t-) = h)$. The process $\alpha_{hl}(t)$ is defined as

$$\alpha_{hl}(t) = \sum_{i=1}^{n} \alpha_{i,hl}(t), \tag{1.38}$$

where $\alpha_{i,hl}(t)$, for $i = 1, \ldots, n$, is the intensity process of the transition from the state h to l for individual i. An important fact to underline is that $\alpha_{hl}(t)$, for $h \neq l$, represent the transition intensities defined in (1.34), and they depend on the history $\mathcal{F}_t$ only through their dependence on the current state h.

Statistical inference for a general Markov multi-state model is here neglected, as it will be illustrated in the special case of the competing risks model. For further theory we refer to Andersen et al. (1993, Chap. 4).

1.4.4 Competing risks models

In this section the competing risks model will be formally presented, while the understanding of its usefulness and its practical use will be widely discussed in an applied framework in Chapter 2.

A competing risks model is in the class of Markov multi-state models. We refer to the notation already used in Section 1.4.1. Consider a stochastic process $\{Z(t), t \in [0, \tau]\}$, with right-continuous paths $Z(t+) = Z(t)$, and finite state space $\mathcal{S} = \{0, 1, \ldots, k\}$.

The state 0 is the only transient state, while the remaining states $\{1, \ldots, k\}$ are absorbing. Usually, in an epidemiological context, the absorbing states represent different types of events, or different causes of the event under study is due to. By definition of the states in $\mathcal{S}$, only the transition probabilities from the state 0 to the states $\{1, \ldots, k\}$ are positive, and therefore considered in a competing risks model. Each of these transitions represents the occurrence of the event due to cause h, with $h \in \{1, \ldots, k\}$ (here we choose to indicate the absorbing states and the causes with the same notation). Moreover, given a realization of the process $Z(t)$, only one of these transitions is observed.

The distribution of $Z(t)$ is regulated by the transition matrix

$$P(s,t) = \{P_{hl}(s,t), h, l \in \mathcal{S}\},$$

where the positive transition probabilities are $P_{0h}(s,t) = P(Z(t) = h \,|\, Z(s) = 0)$, with $h \in \{1, \ldots, k\}$ and $s < t$. The transition intensities are called cause-specific

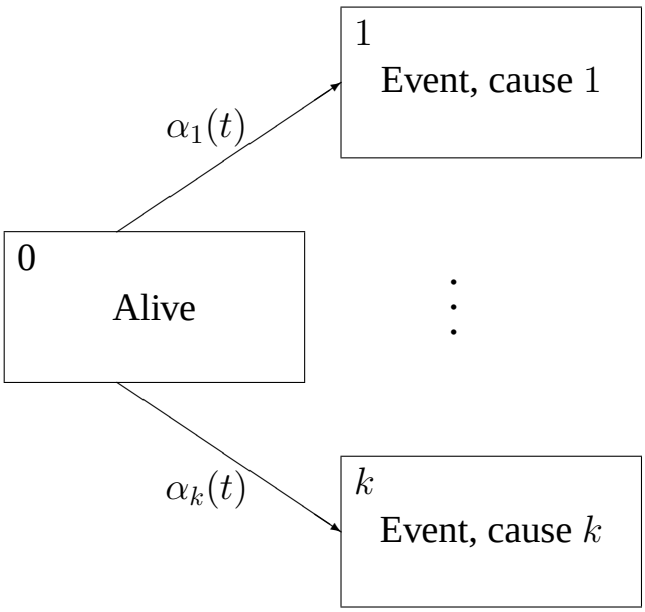

Figure 1.1: Competing risks model with k different causes of the event under study.

hazard functions and equal to

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T^* < t + \Delta t, Z(T^*) = h | T^* \geq t)}{\Delta t}, \quad h = 1, \ldots, k, \quad (1.39)$$

where T^* denote the survival time. Note that here the definition is simply $\alpha_h(t) = \alpha_{0h}(t)$.

Therefore, the competing risks model is built by specifying all the cause-specific hazard functions, and it is represented by the diagram shown in Figure 1.1. A useful reference for understanding the model, its applications and related problems is Andersen et al. (2002).

It is important to stress an aspect concerning the terminology used in the competing risks setting: The name ‘risk’ should refer to a transition probability, hence the phrasing ‘competing risks’ referring to the competing probabilities of experiencing different causes of the event, while the name ‘rate’ is appropriate for describing the cause-specific hazard.

A very informative way to illustrate the behaviour of competing risks over time is by

the transition probabilities from the state 0 to the remaining absorbing states. Each of these functions is called the cumulative incidence probability for cause h, or also cumulative incidence function, and its explicit expression is

$$P_{0h}(t) = \int_0^t S(u-)\alpha_h(u)du, \quad h = 1, \ldots, k, \tag{1.40}$$

where the marginal survival function $S(t) = P(T > t)$ is

$$S(t) = P_{00}(0, t) = \exp(-\sum_{h=1}^{k} \int_0^t \alpha_h(s)ds). \tag{1.41}$$

$P_{0h}(t)$ represents the probability of experiencing the event from cause h before time t and it depends on all the cause-specific hazards $\alpha_1(t), \ldots, \alpha_k(t)$ through the survival function $S(t)$ in (1.41).

A regression competing risks model can be constructed via different regression models for the cause-specific hazards. Regression models for two cause-specific hazards can have different and common covariates, as well as different or common parameters.

1.4.5 Counting process representation for competing risks

The model introduced above can be described by the multivariate counting process $N = \{N_h(t), h = 1, \ldots, k\}$ where

$$N_h(t) = I(T^* \leq t, Z(T^*) = h)$$

counts the number of observed failures from cause h. The associated intensity $\lambda(t)$ has components $\lambda_h(t) = Y(t)\alpha_h(t)$ for $h = 1, \ldots, k$, where $\alpha_h(t)$ is the cause-specific hazard defined in (1.39).

Note that for a general multi-state model the risk indicator was previously defined as $Y_h(t) = I(Z(t-) = h)$, for $h \in \{\ldots, k\}$. In the competing risks model instead, the risk indicator $Y(t)$ does not depend on any cause (or, equivalently, any state of the process $Z(t)$), since 0 is the only state where an individual can be at risk and the initial distribution of the Markov process $Z(t)$ is $\pi_0(0) = 1$.

From the martingale decomposition we obtain that the k-dimensional martingale can be written as $M(t) = N(t) - \Lambda(t)$, where the k-dimensional cumulative intensity process $\Lambda(t) = \{\Lambda_h(t) = \int_0^t \lambda_h(u)du, h = 1, \ldots, k\}$ is the compensator of N.

1.4.6 Statistical inference for the competing risks model

For inferential purposes, the attention is concentrated on the cumulative incidence probabilities defined in (1.40). There exist different approaches for estimating the cumulative incidence probability. The standard approach, which will be explained later in this section, consists of estimating all the cause-specific hazards and then combining them in order to estimate $P_{0h}(0,t)$ (Aalen and Johansen, 1978, Fleming, 1978a,b, Andersen et al., 1993). Regression analysis for competing risks data is performed by constructing a single regression model for each cause-specific hazard function. Therefore, estimates of the cause-specific hazards depend on the estimated regression coefficients of each cause-specific model.

Alternative approaches, which are not treated here, attempt to specify directly a regression model for the cumulative incidence probability. They are based on the so-called subdistribution hazards (Gray, 1988, Fine and Gray, 1999, Scheike and Zhang, 2004, Fine, 2001, Scheike et al., 2007). A parallel approach was proposed by Andersen et al. (2003).

In order to describe the standard approach for estimating P_{0h}, we need to extend the notation of Section 1.4.5 to the case of an n i.i.d. sample of k-dimensional counting processes $(N_{i1},\ldots,N_{ik})$. For this purpose we refer to the general counting process notation for a multi-state model illustrated in Section 1.4.3. Similarly to equations (1.36) and (1.37), we denote here

$$N_{\bullet h}(t) = \sum_{i=1}^{n} N_{ih}(t), \tag{1.42}$$

with $N_{ih} = N_{i,0h}$, and

$$Y_{\bullet}(t) = \sum_{i=1}^{n} Y_i(t),$$

where $Y_i(t)$ is the at-risk indicator for individual i.

We then have the decomposition

$$dN_{ih}(t) = Y_i(t)dA_h(t) + dM_{ih}(t), \tag{1.43}$$

where $A_h = \int \alpha_h(u)du$, for $h = 1,\ldots,k$, are the cause-specific cumulative hazard functions, and M_{ih} is the martingale associated with N_{ih}. Formula (1.43) suggests that

a natural estimator for A_h is

$$\hat{A}_h(t) = \int_0^t \frac{J(s)}{Y_\bullet(s)} dN_{\bullet h}(s), \tag{1.44}$$

where $J(t) = I(Y_\bullet(t) > 0)$ with the convention that $0/0 = 0$. The Nelson-Aalen type estimators in formula (1.44) are determined as the solution to estimating equation based on martingales (Aalen, 1975, 1978). These estimators can also be obtained as nonparametric maximum likelihood estimators (Andersen et al., 1993, Chap. 4).

We now return to the cumulative incidence probabilities, which can be synthesized into the transition probability matrix $P(0,t)$ of the Markov process. Since we are in a competing risks setting, the matrix $P(0,t)$ has elements $P_{0h} \in (0,1)$, for $h = 0, 1, \ldots, k$, on the first row, $P_{hh} = 1$ for $h = 1, \ldots, k$, and zero otherwise. Using the product integral representation in (1.35), the transition matrix can then be written as

$$P(0,t) = \prod_{u \in (0,t]} \left(I + dA(u)\right), \tag{1.45}$$

where A is the matrix of cause-specific cumulative hazard functions. The first row in A is $(A_{00}, A_1, \ldots, A_h, \ldots, A_k)$, with $A_{00} = -\sum_{h=1}^k A_h$, and all the other entries are zero. Relation (1.45) suggests that an estimator of $P(0,t)$ is

$$\hat{P}(0,t) = \prod_{u \in (0,t]} \left(I + d\hat{A}(u)\right), \tag{1.46}$$

where $\hat{A}$ is the matrix constructed from the Nelson-Aalen estimators $\hat{A}_h$ given in (1.44). The first row in $\hat{A}$ is $(\hat{A}_{00}, \hat{A}_1, \ldots, \hat{A}_h, \ldots, \hat{A}_k)$, with $\hat{A}_{00} = -\sum_{h=1}^k \hat{A}_h$, and all the other entries are zero. The estimator in (1.46) is referred to as a product limit estimator of $P(0,t)$, and it is generally called the Aalen-Johansen estimator (Aalen and Johansen, 1978). For a discussion on the variance of this estimator we refer to Martinussen and Scheike (2006, Chap. 10) and Andersen et al. (1993, Chap. 4).

Maximum likelihood estimation under the competing risks model can be based on the likelihood function for the multivariate counting process $N = (N_1, \ldots, N_k)$ with intensity process $\lambda = (\lambda_1, \ldots, \lambda_k)$, as in Section 1.4.5. We consider the intensity process $\lambda^\theta(t)$ depending on a parameter θ, but in order to simplify the notation, we write simply $\lambda(t)$. In the case of complete observations, the likelihood function for θ

up to time t is proportional to

$$L(\theta, t) = \prod_h \prod_{s \leq t} (\lambda_h(s))^{dN_{\bullet h}(s)} \exp\left\{ - \int_0^t \lambda_\bullet(s) ds \right\}, \tag{1.47}$$

where $\lambda_\bullet(t) = \sum_{h=1}^k \lambda_h(t)$, and $dN_{\bullet h} = \sum dN_{ih}$. The log-likelihood function is then equal to

$$l(\theta, t) = \sum_h \left[\int_0^t \log(\lambda_h(s))\, dN_{\bullet h}(s) - \int_0^t \lambda_h(s)\, ds \right].$$

When we have incomplete observations, and we assume independent right-censoring, then the likelihood function in (1.47) is referred to as a partial likelihood function. A general expression for the partial likelihood on the entire interval $[0, \tau]$ is

$$L(\theta) = \prod_{i=1}^n \exp\left\{ - \int_0^{T_i} \lambda_\bullet(s) ds \right\} \prod_{h=1}^k \lambda_h(T_i)^{I(Z_i(T_i)=h)}.$$

In this case the full likelihood would also contain terms corresponding to the distribution of the censoring times.

Chapter 2

Competing Risks Modelling for Breast Cancer Chemotherapy

The competing risks modelling offers a rich set of solutions for several practical problems in biostatistics. In this chapter we shall present a regression model for the competing risks analysis of patients treated for advanced breast cancer. The aim of the application is to detect the optimal chemotherapy dosage for different typologies of patients with advanced breast cancer in order to control the risk of cardiotoxicity. The conditional cumulative incidence probability of getting cardiotoxicity is estimated as a function of the time-dependent covariate 'dosage'. We shall describe the standard approach for studying competing risks and the problems related to its enforceability. We shall also show problems, difficulties and some proposals about how to handle time-dependent covariates.

The application and its results presented in this chapter are based on the paper Ryberg et al. (2008).

2.1 Introduction and background

In the framework of multi-state models (Andersen and Keiding, 2002), explained in Section 1.4 of the thesis, a competing risks model has a transient state, called '0', and a certain number k of absorbing states. Transitions from the state '0' to the k ending states, each one representing the event happening from cause h, with $h = 1, \ldots, k$, are regulated by the cause-specific hazards $\alpha_1(t), \ldots, \alpha_h(t), \ldots, \alpha_k(t)$. As the likelihood function for right-censored survival times depends on these cause-specific hazards,

models based on competing risks can be formulated by their specification. Thus, inferential procedures are straightforward and estimators related to the classical hazard models described in Section 1.3.5 can also be applied in the competing risks models.

2.1.1 The standard approach for regression analysis of competing risks

When the purpose of the study is to investigate the dependence of each transition probability on possible covariates, a regression analysis of competing risks data is required. In this section, we describe the standard approach (Andersen et al., 2002) for a basic competing risks model, within the framework of multi-state models. This methodology is based on the general background presented in Section 1.4. The approach consists of simple regression models for each cause-specific hazard, which are then combined together through the transition probabilities of a well-defined random process. For a competing risks model the process is Markovian, i.e., the cause-specific hazards $\alpha_h(t)$, for $h = 1, \ldots, k$, depend only on the state occupied by the process at time t. In analysing the model related to cause h, where the failure time is due to the cause h, failures due to the competing causes are treated as censored observations.

The attention is concentrated on the cumulative incidence probabilities for the cause-specific events in the time period $(0, t]$. From equation (1.40), these transition probabilities can be written as

$$P_{0h}(0, t; X) = \int_0^t S(u-; X)\alpha_h(u; X)du, \qquad h = 1, \ldots, k, \tag{2.1}$$

where

$$S(t; X) = P_{00}(0, t; X) = \exp\left(-\sum_{h=1}^{k} \int_0^t \alpha_h(u; X)du\right) \tag{2.2}$$

is the marginal survival probability as in (1.41). Conditioning on X underlines the dependence of the transition probabilities on the covariates through the cause-specific hazards, on which specific regression models are built. In $S(t; X)$, the addends $A_h(t; X) = \int_0^t \alpha_h(u; X)du$, for $h = 1, \ldots, k$, are the cumulative cause-specific hazards from the regression models.

Given the observed data, the cumulative incidence probability $P_{0h}(0, t; X)$ is estimated by the Aalen-Johansen type estimator (Aalen and Johansen, 1978, Borgan, 1998), derived from the product-integral representation for the transition probability matrix of the Markov process under study (see equation (1.46)). In order to compute easily the Aalen-Johansen type estimator, it is necessary first to estimate each cumula-

tive cause-specific hazard through a regression model.

Let $\hat{\beta}_h$, $h = 1, \ldots, k$, be the vectors of parameter estimates in the regression models for the cause-specific hazards. Calling X_h the covariate vector in the h cause-specific hazard model, for each $h = 1, \ldots, k$, the estimated increments of the k cumulative cause-specific hazards can be written as

$$d\hat{A}_h(t; X_h) = d\hat{A}_{h0}(t, \hat{\beta}_h)\exp\left\{X_h^T \hat{\beta}_h\right\}, \quad h = 1, \ldots, k. \tag{2.3}$$

Note that some of the covariates considered in the regression analyses can be common to different specific models.

We denote by $t_{h1}, \ldots, t_{hj}, \ldots, t_{hK_h}$ the times when events of type h are observed, for $h = 1, \ldots, k$. K_h is the total number of observed events of type h. Then, $\hat{A}_{h0}(t, \hat{\beta}_h)$ for $h = 1, \ldots, k$ are the Breslow estimators of the baseline cumulative cause-specific hazards, i.e.,

$$\hat{A}_{h0}(t, \hat{\beta}_h) = \sum_{t_{hj} \leq t} \frac{dN_{\bullet h}(t_{hj})}{S_h^{(0)}(t_{hj}, \hat{\beta}_h)}, \quad h = 1, \ldots, k, \tag{2.4}$$

where $dN_{\bullet h}(t_{hj})$, defined in (1.42), is the number of events of type h which occurred at time t_{hj}. The quantity $S_h^{(0)}(t, \hat{\beta}_h)$ is defined as

$$S_h^{(0)}(t, \hat{\beta}_h) = \sum_{i=1}^{n} \exp\left\{X_{h,i}^T \hat{\beta}_h\right\} Y_i(t), \quad h = 1, \ldots, k,$$

with $Y_i(t)$ the indicator for patient i at risk at time $t-$, and $X_{h,i}$ contains the covariate values for individual i.

In the competing risks model, in order to estimate the probability P_{00} of being alive at time t without experiencing any event due to the k causes, we can use the the Kaplan-Meier type estimator (Kaplan and Meier, 1958)

$$\hat{P}_{00}(0, t; X_h, h = 1, \ldots, k) = \prod_{t_j \leq t} \left\{1 - \sum_{h=1}^{k} d\hat{A}_h(t_j; X_h)\right\}. \tag{2.5}$$

It is computed using the estimated increments $\hat{A}_h$ of the cumulative cause-specific hazards, given in (2.4), for $h = 1, \ldots, k$. It is important to note that the product is computed at times t_j, which are times when an event of any type happens.

In the competing risks setting the interest is often addressed to the study of one specific of the k possible causes, let it be h, and to the corresponding cumulative incidence probability. The cumulative incidence probability for the cause-h event can then be

estimated by the Aalen-Johansen type estimator, by plug-in of all the k estimated increments of the cumulative cause-specific hazards and the estimated marginal survival probability:

$$\hat{P}_{0h}(0,t;X) = \sum_{0<t_{hj}\leq t} \hat{P}_{00}\left(0,t_{hj}-;X\right) \mathrm{d}\hat{A}_h(t_{hj};X_h). \tag{2.6}$$

We underline that this estimator works correctly in case all the covariates are time-independent in the considered time interval $(0,t]$. A time-dependent covariate can also be included when it is already defined at the time origin, since it can then be considered a deterministic path (Kalbfleisch and Prentice, 2002, Chap. 5).

2.1.2 Residuals for goodness-of-fit of the cause-specific hazard models

In case the standard approach in subsection 2.1.1 is applied, the goodness-of-fit for the competing risk model relies on the diagnostics for each cause-specific hazard model.

In this section we limit our attention on the basic ideas for the residuals in the Cox model (Example 1.3.2), even though they are also used elsewhere. We briefly summarize the theory concerning the different types of residuals used later on in the current chapter within the application to breast cancer. A general reference for this section is Therneau and Grambsch (2000, Chap. 4).

Martingale residuals

The martingale process for the individual i, in case of a Cox model, is written as

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp\left\{X_i^T(s)\beta\right\}\lambda_0(s)\mathrm{d}s, \tag{2.7}$$

where $X_i(\cdot)$ is the possibly time-dependent p-dimensional covariate vector for individual i. The martingale residual process is then defined as

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\exp\left\{X_i^T(s)\hat{\beta}\right\}\mathrm{d}\hat{\Lambda}_0(s), \tag{2.8}$$

where $\hat{\Lambda}_0(t)$ estimates the cumulative baseline intensity process and $\hat{\beta}$ is the vector of the maximum partial likelihood estimates of the regression coefficients. The term $\hat{\Lambda}_i(t) = \int_0^t Y_i(s)\exp\left\{X_i^T(s)\hat{\beta}\right\}\mathrm{d}\hat{\Lambda}_0(s)$ estimates the compensator of the counting process $N_i(t)$ and represents the estimated cumulative intensity for individual i in the Cox model.

The martingale residuals are defined as the martingale residual processes at the end of the study. Formally they are

$$\hat{M}_i(\infty) = N_i(\infty) - \hat{\Lambda}_i(\infty), \qquad \text{for } i = 1, \ldots, n. \tag{2.9}$$

We can write them more synthetically as $\hat{M}_i = N_i - \hat{\Lambda}_i$.

When the hazard function follows a Cox model, in case of no time-dependent covariates, the martingale residual for individual i takes the form

$$\hat{M}_i = N_i - \exp\left\{X_i^T \hat{\beta}\right\} \int_0^\infty Y_i(t) \mathrm{d}\hat{\Lambda}_0(t), \tag{2.10}$$

where X_i is the covariate vector for individual i. It represents the difference between the observed number of events for subject i (N_i) and the expected number of events conditional to the observed data.

The following properties are essential to understand the practical use of these residuals:

- $\sum_{i=1}^n \hat{M}_i = 0$, i.e. the sum of martingale residuals, given the estimate $\hat{\beta}$, is equal to zero;
- $E(\hat{M}_i) = 0$, i.e. the expected value of each residual is equal to zero at the true parameter β;
- $\mathrm{Cov}(\hat{M}_i, \hat{M}_j) = 0$ for $i \neq j$, i.e. the residuals at the true parameter β are uncorrelated.

Schoenfeld residuals (Schoenfeld, 1982)

We consider the p-dimensional score process over subjects

$$U(\beta, t) = \sum_{i=1}^n \int_0^t \{X_i(s) - E(\beta, s)\} \, \mathrm{d}M_i(s), \tag{2.11}$$

where $E(\beta, s) = S_1(\beta, t)/S_0(\beta, t)$, with $S_0(\beta, t) = \sum_{i=1}^n Y_i(t) \exp\left\{X_i^T(t)\beta\right\}$ and $S_1(\beta, t) = \sum_{i=1}^n Y_i(t) X_i(t) \exp\left\{X_i^T(t)\beta\right\}$. From the estimating equation for the Cox model, the observed score process is then equivalent to

$$U(\hat{\beta}, t) = \sum_{i=1}^n U_i(\hat{\beta}, t) = \sum_i \int_0^t \left\{X_i(s) - E(\hat{\beta}, s)\right\} \mathrm{d}\hat{M}_i(s).$$

From the martingale residual decomposition $\mathrm{d}\hat{M}_i(t) = \mathrm{d}N_i(t) - \mathrm{d}\hat{\Lambda}_i(t)$ and the ex-

pression of the estimator of the cumulative baseline hazard, it is easy to verify that, when there are no ties for the failure times, it is also

$$U(\hat{\beta}, t) = \sum_i \int_0^t \Big\{ X_i(s) - E(\hat{\beta}, s) \Big\} \, \mathrm{d}N_i(s).$$

The observed score process is piecewise constant over time and it has jumps in correspondence with failure times. Therefore, Schoenfeld residuals come from the idea of splitting $U(\hat{\beta}, \infty)$, the observed score process at the end of the study, in the time intervals identified by the failure times $t_1, \ldots, t_k, \ldots$ observed from the data. The Schoenfeld residual at the failure time t_k is then

$$s_k = \int_{t_{k-1}}^{t_k} \sum_i \Big\{ X_i(s) - E(\hat{\beta}, s) \Big\} \, \mathrm{d}N_i(s). \tag{2.12}$$

The residual s_k in (2.12) is really a p-dimensional vector, because $U_i(\hat{\beta}, t)$ is a p-dimensional process with components $U_{ij}(\hat{\beta}, t) = \int_0^t \{X_{ij}(s) - E(\hat{\beta}, s)\} \mathrm{d}N_i(s)$ for the covariates $j = 1, \ldots, p$.

Notice that Schoenfeld residuals in (2.12) are still valid in case of a time-dependent covariate. In case of no ties, their computation is easy and can be based on the following expression

$$s_k = X_{(k)}(t_k) - E(\hat{\beta}, t_k)$$

for each failure time t_k, where $X_{(k)}(t_k)$ is the p-dimensional covariate vector at time t_k for the subject who experienced a failure at t_k.

Cumulative residuals

Cumulative residuals (Lin et al., 1993, Wei, 1984) are used to test various assumptions in the Cox model, such as the functional form of covariates, the proportionality of the hazards and misspecification of the link function. They are therefore expressed as different functionals of the residuals $\hat{M}_i(t)$ in (2.8).

The simplest functional of residuals is the observed score function depending on time,

$$U(\hat{\beta}, t) = \sum_{i=1}^{n} \int_0^t X_i(s) \mathrm{d}\hat{M}_i(s).$$

The cumulative martingale residuals, given by each component $U_j(\hat{\beta}, t)$ for $j = 1, \ldots, p$, are useful in checking the proportional hazards assumption.

Other types of cumulative residuals are obtained by considering a two-dimensional cumulative residual process depending both on time and on covariate values (Lin et al., 1993, 2000), or by partial sums of the estimated residuals depending only on the covariate values as illustrated later in Section 2.3.3.

2.2 An application to breast cancer: Introduction and scope of the study

Breast cancer has become a major health problem over the last 50 years and worldwide, it is the fifth most common cause of cancer death. The anthracycline based on Epirubicin is among the most commonly used antitumour chemotherapy with activity against a wide spectrum of cancer diseases (Tormeys, 1975, Goldin et al., 1985). Nevertheless, it was demonstrated that its antitumour effect is set off against its cardiotoxic side effects such as cardiomyopathy and congestive heart failure (CHF)(Brambilla et al., 1986). The risk of cardiotoxicity after anthracycline-based treatment has been shown to depend on the cumulative dose administered to patients and seems to increase in case of some risk factors such as preexisting cardiac disease or previous irradiation against the heart (Swain et al., 2003). As the median survival for patients with advanced breast cancer is short, a 5% risk for development of CHF is generally accepted and it is estimated to correspond to a total dose of 950 mg/m^2 of Epirubicin (Ryberg et al., 1998). In previous medical studies, this recommended cumulative dose was determined by the Kaplan-Meier estimator as a function of the cumulative dose only (Swain et al., 2003, Ryberg et al., 1998). Thus, this statistical analysis ignored both the effect of time and the competing risk of dying from advanced cancer. The application of the competing risks method to this problem, presented in this chapter, has compensated for the missing aspects, taking both the cardiotoxicity and mortality rates into account.

The general scope of the study was to assess an optimal recommended total dose taking the following aspects into account: history of the dose administration during the treatment period, concurrent risk of dying of advanced cancer and possible predictors for cardiotoxicity. Investigation on possible predictors for development of CHF has been of primary interest because they have allowed discriminating between recommended cumulative doses for different groups of patients.

A well-defined cohort of 1097 patients treated with an Epirubicin based chemotherapy for advanced breast cancer admitted to Herlev Hospital (Denmark) has been retrospec-

Treatment time schedule	
Days of administration and doses per day	*Frequencies*
Epi day 1 and day 8: 60 mg/m^2 every 4 weeks	70
Epi day 1 and day 8: 70 mg/m^2 every 4 weeks	181
Epi day 1: 130 mg/m^2 every 3 weeks	164
Epi day 1: 130 mg/m^2 every 6 weeks + CTX	49
Epi day 1: 100 mg/m^2 every 3 weeks	514
Epi day1 and day 8: 45 mg/m^2 every 4 weeks + Vindesine	54
Epi day1 and day 8: 65 mg/m^2 every 4 weeks + Cisplatin	65
Treatments	
Type	*Frequencies*
Epirubicin	929
Epirubicin + Vindecine	54
Epirubicin + Cisplatin	65
Epirubicin + CTX	49

Table 2.1: The complete treatment based on Epirubicin. Some groups underwent additional chemotherapy (CTX, Vindesine, Cisplatin). The time schedule varies according to days of administration and doses per day.

tively analysed during a period of twenty years (from November 1983 to November 2003). The patients had no evidence of cardiac disease or a history of myocardial infarction before starting the chemotherapeutic treatment. The women in the study followed different treatment regimes for Epirubicin. Some of them received an additional chemotherapy. Information about type of treatments and time schedule are shown in Table 2.1. However, because of the seriousness of the advanced cancer stage and the collateral symptoms, almost all patients deviated from their schedule.

Hereafter we explain the aims of the present study in greater details and the overview of the corresponding statistical methods appropriately applied. In the next section, these statistical methods together with problems related to their applicability shall be discussed, placed side by side with numerical results.

The first aim of the study was to investigate which predictors were significant for developing CHF. The second aim was to estimate the conditional probability of developing CHF within a certain time interval, as a function of epirubicin cumulative dose (and of other prognostic factors), taking also the possibility of dying from breast cancer into account. In order to evaluate this probability, a competing risks model with two causes was suitable. Cardiotoxicity was the event of primary interest and mortality from breast cancer was the competing event. Figure 2.1 shows graphically the competing risks model as a multi-state model with two possible ending events. In

the framework of competing risks, the possibilities for a patient along the follow-up time are to be alive with no sign of CHF, to develop cardiotoxicity and to die from breast cancer without cardiotoxicity. These three states must be taken into consideration when a patient's risk for cardiotoxicity is estimated, as well as when the risk of dying from breast cancer is evaluated. Follow-up time was from start of the epirubicin treatment until patients either developed cardiotoxicity, died without cardiotoxicity or left the study alive without cardiotoxicity.

The competing risk model depends on both the cause-specific hazards $\alpha_c(t)$ and $\alpha_d(t)$, as observed from Figure 2.1. Thus, as the probability of developing CHF within a certain time interval depends on cumulative dose through both the mortality rate from breast cancer and the cardiotoxicity rate, the statistical analysis was carried out in two steps. In the first one (corresponding to the first aim), both of the two competing event rates were estimated through regression analyses, considering the possible effect of cumulative dose and other prognostic factors. In the second step (corresponding to the second aim), the estimated rates were used to evaluate the cumulative incidence probability for CHF, i.e. the probability of developing CHF within a certain time interval. Finally, when this probability was fixed equal to 5%, it was possible to determine the corresponding value of the cumulative dose in order to find the optimal recommended total dosage.

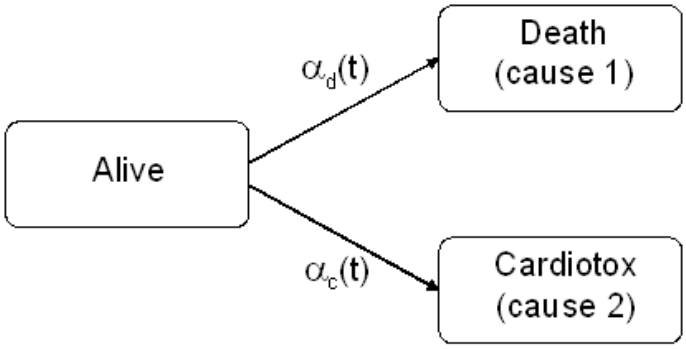

Figure 2.1: *Competing risks model with two causes*

2.3 The regression models for the cause-specific hazards

The influence of the Epirubicin cumulative doses along the treatment period and other prognostic factors on both the cardiotoxicity rate and the mortality rate from breast cancer, was investigated by Cox regression models (Cox, 1972) (see Example 1.3.2 in Section 1.3.5) separately for each one of the two rates. When modelling the cardiotox-

icity rate (i.e. development of CHF), the event time was considered right-censored for patients who died without having cardiotoxicity. On the contrary, when modeling the mortality rate, the event of interest was death without cardiotoxicity; therefore, event times for the patients who developed CHF were right-censored.

The choice of distinct regression parameters for the covariates in the two regression models was due to the fact that covariates are expected to influence the two rates differently.

2.3.1 Assumptions and preparation of the data set

The preparation of the data was nontrivial and computer expensive, but essential, especially for treating the cumulative treatment dose as a time-dependent covariate in the regression analyses.

In the data set the information for each patient consisted in the number of dose injections, date of the last injection, total dose administrated, date of entry in the study, treatment duration, presence of cardiotoxicity and the possible date of its development, date of death or last seen.

Patients followed different treatment schedules (Table 2.1) and, as mentioned in Section 2.2, each woman had deviation with respect to the dose schedule, but the exact information about the single doses was not available. For this reason, we assumed that the patient was given the same dose amount at each injection, calculated as her total dose divided by the number of injections. Moreover, each individual had its own treatment period and her time schedule could also deviate from the protocol because of missing or additional injections, but information about the exact times of injections was not available. Therefore, different functions corresponding to the three different time schedules, were implemented in order to calculate the assumed dates of injections for each patient. As an example, let us explain the computation of the dates of dose administration for a patient who should have followed the time schedule 'day 1 and day 8 every 4 weeks'. Figure 2.2 can simplify the understanding of the elaboration. We define r_i and d_i the number of injections and the total treatment duration of patient i, respectively. If the patient was following exactly the right time schedule, the intervals $I_i^{(l)}$ and $I_i^{(s)}$ in Figure 2.2 should have been of length 4 weeks and 8 days, respectively. In our situation doctors were unable to follow the protocol in all the cases. Therefore, we decided to find the lengths of the intervals $I_i^{(l)}$ and $I_i^{(s)}$ by keeping the ratio $I_i^{(s)}/I_i^{(l)}$ equal to the one of the schedule (8/28). In case of $r_i > 2$, if r_i was even,

the function $(d_i - 8)/\left[(r_i/2) - 1\right]$ was used to assume the length of $I_i^{(l)}$, while if r_i was odd, the function $d_i/\left[(r_i - 1)/2\right]$ was computed. Length of $I_i^{(s)}$ was obtained by $I_i^{(l)} \cdot 8/28$. In case r_i was equal to 2, the length of $I_i^{(s)}$ was set equal to d_i and $I_i^{(l)}$ was empty.

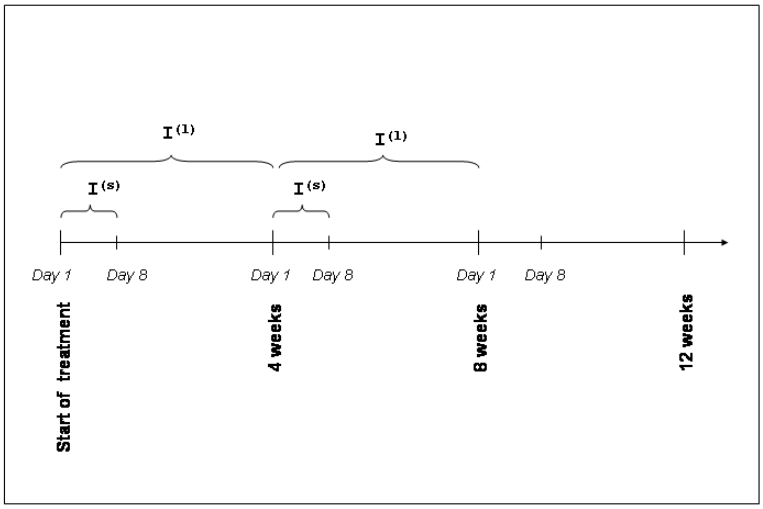

Figure 2.2: *The interval $I^{(s)}$ between two single dose injections and the interval $I^{(l)}$ of the first treatment cycle for a patient how followed exactly the time schedule 'day 1 and day 8 every 4 weeks'.*

These last computations led to represent the information of each patient about Epirubicin dosage by a vector containing history of cumulative doses at every time of administration from the date of entry in the study to the date of end of treatment, event or censoring. Note that in our data the treatment period can be shorter or equal to the follow-up period. Thus, cumulative dose of epirubicin (mg/m^2) was considered as a time-dependent variable in both the regression models. Statistical analysis in case of time-dependent covariates needs a special formulation of the data. The counting process form is a very useful instrument to represent information on these covariates in the dataset. The entire follow-up period of a subject is split in risk intervals, which are built on the time points where the covariate changes its value. In our application, the cumulative dose is an increasing step function, with jumps at the end of each risk interval. Each subject is then represented by a set of observations, one for each risk interval. Intervals are defined as left-open so that covariate history is a predictable process and the event or censoring coincides with the upper bound of the interval, as for the risk indicator in the counting process notation.

The dataset was adjusted for presence of ties (Therneau and Grambsch, 2000, Chap. 2) and overlapping of date of event and date of ending treatment. When the latter case

happened, dates were translated by 0.5 days. This was done in order to be able to evaluate information about the time-dependent dose at the end of the treatment in the Cox models.

Covariates involved in the study and measured at start of treatment were age, performance status (PS), number of sites affected by the tumour, type of complete treatment (single drug or additional treatments), predisposition to cardiac diseases, previous treatment for breast cancer either in an adjuvant setting or for relapse (antihormonal therapy, or chemotherapy), adjuvant or extensive radiotherapy and palliation radiotherapy (local skin metastases, thoracic spine, mediastinum).

2.3.2 The Cox regression models

Out of 1097, 125 patients developed CHF while in 10 patients the information about the CHF development was either missing or uncertain. The number of patients who died was equal to 1032. There was presence of missing values for some covariates.

Care needs to be taken when information about the event due to a certain cause h is missing for some patients. In this case, estimates in all the cause-specific regression models would need to be computed ignoring patients with missing information on the event from cause h. Otherwise, we may overestimate the observed number of events from causes other than h, in studies where patients can experience multiple events. In our study, we ignored 10 patients with missing information about the CHF development in both the Cox regression models.

We study the problem under the independent right-censoring assumption. In order to investigate the best statistical model for the cardiotoxicity rate, detailed analyses were performed to evaluate, in the following order, the appropriate functional form for the continuous variables, which covariates were significant risk factors, whether the proportional hazards assumption in the Cox model was correct and a possible stratification. The same analyses were also performed for finding a correct regression model for the mortality rate for breast cancer, taking presence of cardiotoxicity into account by censoring.

The correct functional form was investigated for the covariates age and cumulative dose. For testing the correct functional form, graphical methods based on martingale residuals and smoothing splines were used (Therneau et al., 1990). In both the Cox models for $\alpha_c(t)$ and $\alpha_d(t)$, results showed that no transformations of the functional form of the cumulative dose and age were necessary.

Separate procedures based on backward and forward algorithms were used to select the group of covariates significant at the 5% level. Interactions between each covariate and the cumulative dose were also considered when applying these procedures. Likelihood ratio tests were performed in order to decide whether to include the interaction term between a covariate and the cumulative dose into the model, besides the single effect of the covariate.

The proportionality of the hazards was investigated by graphical methods and by tests of hypothesis based on Schoenfeld residuals (Grambsch and Therneau, 1994). Presence of non-constant coefficients in the model indicates that the effect of some covariates on the hazard may vary over time, thus violating the assumption of proportionality. In the model for the cardiotoxicity rate, the global test and the univariate test of non-proportionality for each covariate were not significant. In the model for the mortality rate, proportionality of the hazard did not hold for performance status, which was then used as a stratification factor. Moreover, since the effect of the cumulative dose was found to vary over time, a time-varying coefficient (Scheike, 2004) was used. Therefore, follow-up was split in three time intervals and the cumulative dose was studied separately in each time interval using different coefficients in the model. The three time intervals were the first three months after the start of treatment, the following three months and from the seventh month onwards.

In the final Cox model, the cardiotoxicity hazard function for individual i was:

$$\alpha_{c,i}(t;X_c) = \alpha_{c0}(t)\exp\{X_{c,1i}(t)\beta_1 + X_{c,2i}\beta_2 + X_{c,3i}\beta_3 + X_{c,4i}\beta_4 + X_{c,5i}\beta_5 + X_{c,6i}\beta_6 + (X_{c,1i}(t) * X_{c,6i})\beta_7\},$$

where the unspecified nonnegative function $\alpha_{c0}(t)$ is the baseline hazard for cardiotoxicity and X_c is the covariate vector with components $X_{c,l}$, $l = 1,\ldots,6$, defined to be, respectively, time-dependent cumulative dose of epirubicin, disposition to heart disease, previous antihormonal treatment for relapse, irradiation against thoracic spine, age, and previous chemotherapy (CMF) for relapse. The lth covariate of individual i is specified as $X_{c,li}$, where $l = 1,\ldots,6$. A significant interaction was found between cumulative dose, X_1, and CMF for relapse, X_6.

In the second Cox model, the mortality hazard function for individual i was:

$$\alpha_{d,i}(t;X_d) = \alpha_{d0}^{(k)}(t;X_9)\exp\{X_{d,1i}(t)\beta_j(t) + X_{d,5i}\beta_{11} + X_{d,7i}\beta_{12} + X_{d,8i}\beta_{13}\},$$

where the vector X_d contains the covariates $X_{d,l}$, for $l = 1,5,7,8,9$. Covariates $X_{d,7}$,

$X_{d,8}$ and $X_{d,9}$ are defined to be, respectively, previous adjuvant chemotherapy, presence of more than one tumour site and performance status. Variation of the effect of cumulative dose in time is represented by the different coefficients $\beta_j(t)$, $j = 8, 9, 10$, such that

$$\beta_j(t) = \begin{cases} \beta_8 & \text{if} \quad 0 < t \leq 91.31 \\ \beta_9 & \text{if} \quad 91.31 < t \leq 182.62 \\ \beta_{10} & \text{if} \quad t > 182.62. \end{cases}$$

The stratification factor, X_9 in our application, divides the subjects into disjoint groups, each of which has a distinct baseline hazard function but common values for the regression coefficients. If individual i belongs to stratum (k), then $\alpha_{d0}^{(k)}(t; X_9)$ is her baseline hazard function for mortality.

Results about the two Cox regression models are synthesized in Table 2.2. As in our application the main interest is on the effect of the Epirubicin dose, description of the results focuses on this covariate and it is given separately for the two rates.

The rate of cardiotoxicity was shown to depend log-linearly on the cumulative dose, with different effects for patients with or without CMF for relapse. This difference is due to the significant interaction term between cumulative dose and CMF for relapse. In the group of patients who did not receive CMF for relapse, the cardiotoxicity rate had an increase of 40% each time the cumulative dose increased by 100 mg/m^2, holding the other covariates constant. Thus, from a cumulative dose of 600 mg/m^2 to a level of 900 mg/m^2 the rate was increased 2.72 fold. For the group of patients who received CMF for relapse instead, an increase of 100 mg/m^2 in the cumulative dose was associated with a 91.5% ($\exp\{\beta_1 + \beta_7\}$) increase in the cardiotoxicity rate (Table 2.2). Therefore, presence of CMF for relapse, in addition to its own effect, raised the effect of increasing dose on the CHF rate. This is shown in Figure 2.3, which illustrates the interpretation of the interaction term. The two lines in Figure 2.3 represent the logarithm of cardiotoxicity rate for patients with CMF for relapse and patients without CMF treatment. The slopes of the lines represent the different effects of cumulative dose for the two groups of patients. For doses below 928 mg/m^2, patients who received CMF had a lower cardiotoxicity rate compared to patients without CMF. On the contrary, the picture was inverted at doses higher than 928 mg/m^2. E.g. at an epirubicin dose equal to 950 mg/m^2, the hazard was 7.3% higher for patients with CMF than for patients without CMF. Both the estimated regression coefficients about CMF and its interaction with dose were needed in order to calculate this hazard ($\exp\{\beta_6 + \beta_7(9.5 - 5)\}$). Because of the significant interaction term in the model,

Predictors for the cardiotoxicity rate				
Variable	β	**Robust se(β)**	**Exp (β)**	P**-value**
Cumulative dose (100 mg/m^2)	0.334	0.073	1.396	<0.0001
Disposition to heart disease	1.102	0.209	3.010	<0.0001
Previous antihormonal treatment	0.628	0.214	1.873	0.0033
Irradiation against thoracic spine	0.734	0.251	2.084	0.0035
Age	0.025	0.010	1.025	0.012
CMF for relapse at cumulative dose 500 mg/m^2	-1.350	0.697	0.259	0.053
(CMF for relapse) * (Cumulative dose (100 mg/m^2))	0.316	0.121	1.371	0.0092
Predictors for the mortality rate				
Variable	β	**Robust se(β)**	**Exp (β)**	P**-value**
Cumulative dose (100 mg/m^2) (during the first three months of follow-up)	-1.047	0.164	0.351	<0.0001
Cumulative dose (100 mg/m^2) (during the fourth, fifth and sixth months)	-0.504	0.061	0.604	<0.0001
Cumulative dose (100 mg/m^2) (from the seventh month on)	-0.106	0.016	0.900	<0.0001
Age	0.012	0.004	1.012	0.0026
Adjuvant CMF	0.254	0.078	1.289	0.0011
Number of sites >1	0.721	0.077	2.056	<0.0001

Table 2.2: *Estimates in the Cox regression models for the cardiotoxicity rate and the mortality rate*.

CMF for relapse affected the cardiotoxicity rate, conditionally to the other risk factors, with a magnitude depending on the cumulative dose. The coefficient for CMF, equal to -1.35 (Table 2.2), represents the difference between cardiotoxicity rates for the two groups, when the cumulative dose is fixed at 500 mg/m^2, as shown in Figure 2.3.

The mortality rate was shown to decrease by increasing cumulative dose. When the cumulative dose increased by 100 mg/m^2, the rate was reduced by 65% during the first three months of treatment, by 40% between the fourth and sixth month and by 10% from the seventh month on. Thus, the effect of increasing doses on reducing the mortality rate was higher in the first treatment period than later on. The mortality rate decreased by 82.9% ($\exp\{\beta_9(9.5-6)\}$) going from a cumulative dose of 600 mg/m^2 to a dose of 950 mg/m^2 between the fourth and sixth month of treatment (Table 2.2).

2.3.3 Problems related to goodness-of-fit of regression models

For testing the correct functional form of the continuous covariates age and cumulative dose in the Cox regression models, graphical methods based on martingale residuals and smoothing splines were used (Therneau et al., 1990). In this section we illustrate how these methods were applied and which are the related problems, especially in connection with time-dependent covariates. Methods are first illustrated on the time-constant variable age, and then they are discussed for the time-dependent variable

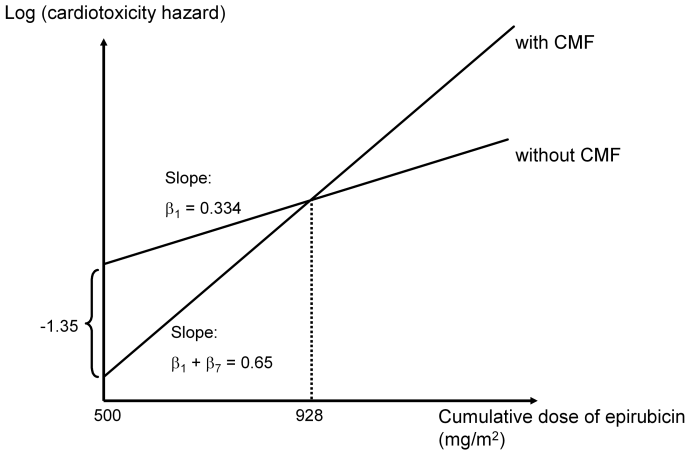

Figure 2.3: *Interpretation of the interaction term between cumulative dose and CMF for relapse in the Cox model for the cardiotoxicity rate. The two lines represent the logarithm of the hazard for patients with CMF for relapse and patients without CMF treatment. The different effects of cumulative dose are represented by the slopes of the lines, which are computed from the estimates of regression parameters.*

cumulative dose, in connection with the proportional hazards model for the cardiotoxicity rate.

For testing graphically if the linear form of a covariate is correct, the simplest approach consists in plotting the logarithm of the estimated cumulative hazard versus the covariate. When investigating the covariate age, the plotted cumulative hazards refer to the simple Cox regression model with the only covariate age and can be estimated from equation (2.3). The resulting graph, together with a scatterplot smooth function, is shown in panel (a) of Figure 2.4 and suggests that the linear form for age might be correct.

The second approach by Therneau et al. (1990), Therneau and Grambsch (2000, Chap. 5) consists in plotting the martingale residuals from a regression model with only the covariate of interest versus the covariate values. In this case, superimposing a smooth function should indicate the correct functional form for the covariate under investigation. As age is a time-independent covariate, martingale residuals are expressed by equation (2.10) when analysing this covariate. Panel (b) in Figure 2.4 shows martingale residuals computed per-individual. If a linear form was correct, we would expect to observe no specific pattern of the data and a linear smooth function superimposed. Nevertheless, interpretation of these concepts for the age graph is hard, as the data points appear to be clustered in an atypical pattern. A large number of residuals are observed between -0.3 and 0, while none is present between 0 and 0.7, few points are higher than 0.7. A possible explanation of this pattern could be the presence of

few events of cardiotoxicity (125) with respect to the censored data. As it was already pointed out by Therneau and Grambsch (2000, Chap. 5), this situation is frequent in data sets with a large amount of censoring. The observed counting process, which is a component of the martingale residuals as observed in (2.9), is equal to 0 for many individuals, and this fact leads to many negative values for the residuals.

An alternative appealing approach by Hastie and Tibshirani (1990, Chap. 2) is to model the functional form of covariates through smoothing splines directly in the Cox model. The use of smoothing splines requires to specify a certain number of knots (degrees of freedom) and therefore results might strongly depend on the chosen degrees of freedom. In order to understand the basic relation between the hazard and a single covariate, a simple Cox model with a smoothing spline for age and no further covariates was considered. Panel (c) in Figure 2.4, where four degrees of freedom were chosen, does not show a significant curvature for the covariate age, as also confirmed by the corresponding test of hypothesis ($p < 0.001$ for the linear term and $p = 0.29$ for the nonlinear term).

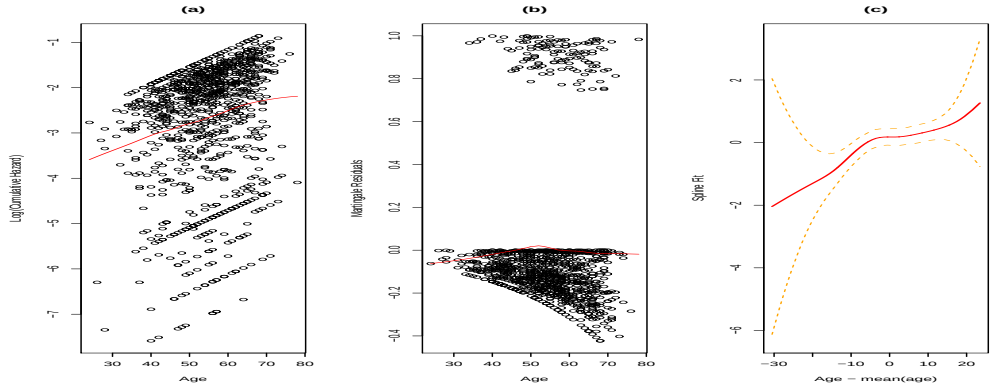

Figure 2.4: *Functional form of the covariate age in the Cox model for the cardiotoxicity rate.*

In the following, the methods applied above to age are illustrated for the covariate cumulative dose, and the differences and critical aspects of investigating the functional form of a time-dependent covariate are discussed. Panel (a) in Figure 2.5 shows the logarithm of the estimated cumulative hazard versus the total cumulative dose. The cumulative hazards are given per-subject and refer to the simple Cox regression model with the only covariate cumulative dose. Estimates are obtained summing the increments in (2.3) up to the entire follow-up period.

The original data set is modified so that each patient is split in more observations and cumulative dose is time-constant in the risk set of each observation. The estimated

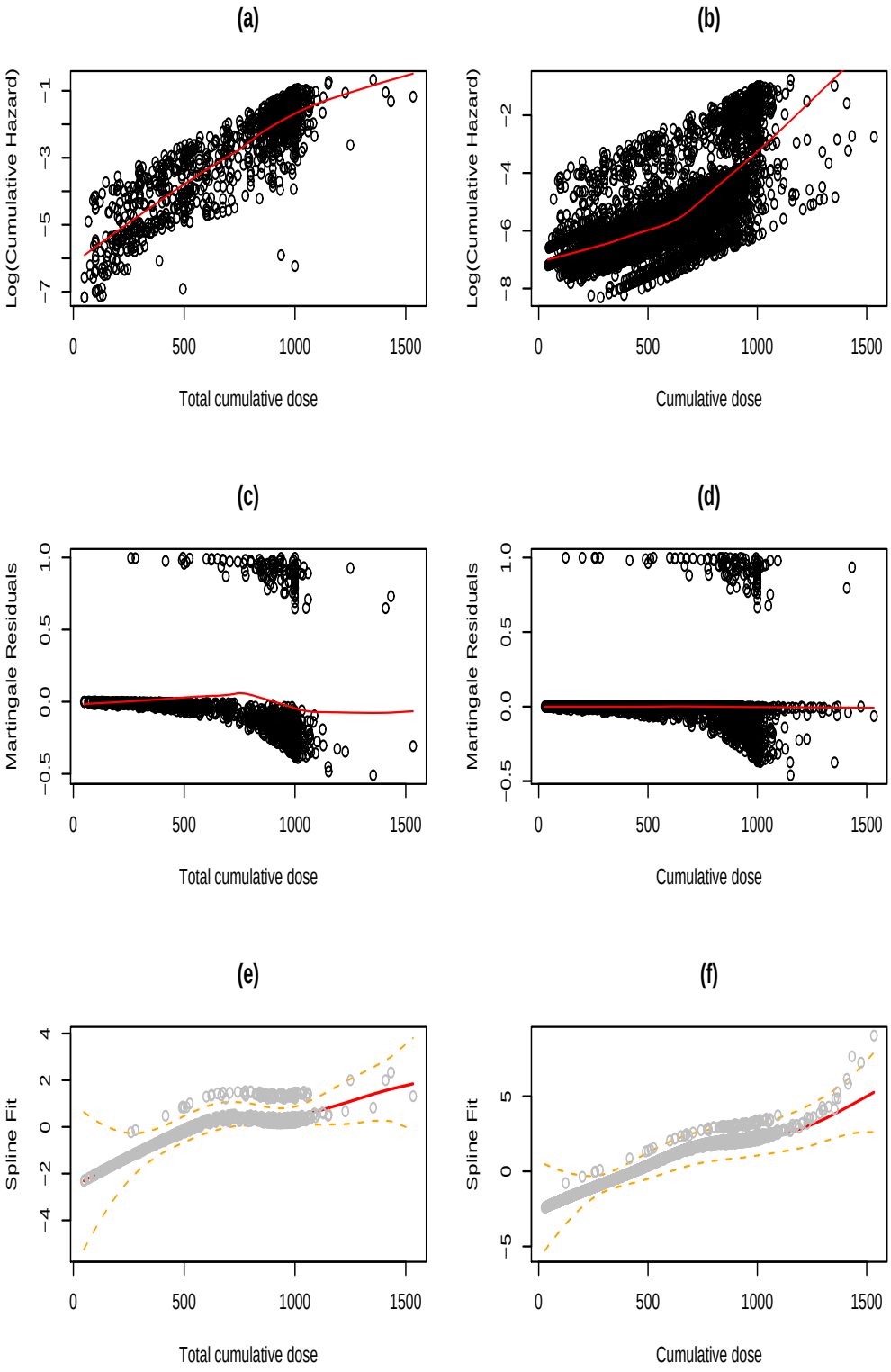

Figure 2.5: *Functional form of the covariate cumulative dose in the Cox model for the cardiotoxicity rate.*

cumulative hazard of each patient is then the sum of the estimated cumulative hazards of her observations. In fact, the per-subject estimated cumulative hazards in panel (a) of Figure 2.5 are obtained by collapsing the estimates related to all the observations that the subject is split in.

The case when the estimated cumulative hazards refer to all the observations of individuals can be observed in panel (b) of Figure 2.5, where a much larger amount of points than in panel (a) is represented. Comparison between panels (a) and (b) might suggest different conclusions about the functional form. This fact might be acceptable as the covariates under investigation are different, the total dose in panel (a) and the cumulative dose in panel (b). The problem is that our interest focuses on the covariate cumulative dose, but investigation of its functional form by computing and plotting per-observation estimates (as in panel (b)) can yield distortions and misinterpretation. The problem is further discussed for the martingale residuals studied later in this section.

The martingale residuals from a regression model with only the covariate cumulative dose are plotted against the covariate values of total dose. Martingale residuals of each patient are given by the sum of martingale residuals of her observations and are shown in panel (c) of Figure 2.5. Formally, martingale residuals of patients in case of a time-dependent covariate, as cumulative dose, can not be computed by equation (2.10). In this equation, the exponential needs to be under the integral. As our interest is on the functional form of the time-dependent covariate cumulative dose, per-observation martingale residuals are also plotted against cumulative dose and the pattern is shown in panel (d) of Figure 2.5. Here, points are clustered in horizontal bands and the interpretation of this pattern in order to check the linear form of the covariate, is quite hard, despite the help from the superimposed smooth function. The same problem occurs in panel (b), where, in comparison with panel (a), an additional band of points is observed. The explanation of these difficulties is that there is a large amount of 'artificial' observations, created in order to handle time-dependent covariates, which leads to many 'artificial' censoring. For this reason, many low values of martingale residuals are observed. Only the latest observation of each individual is eventually associated with a failure, and therefore with a positive residual. Moreover, most of the observations correspond to small or near to zero values of the estimated cumulative hazard, due to the presence of small at risk time intervals, leading then to most of the martingale residuals being nearly zero. Further explanations and more detailed analyses about bias in case of time dependence of covariates are given by Therneau

and Grambsch (2000, Chap. 5).

The approach using smoothing splines, choosing 4 degrees of freedom for the knots (Hastie and Tibshirani, 1990, Chap. 2), is also considered in order to test the correct functional form of the cumulative dose. This approach is not affected by the problems previously discussed about the time-dependent covariate. A spline fit is shown in panel (f) of Figure 2.5 and suggests a linear form for the covariate. This result is also confirmed by the accompanying test of hypothesis ($p < 0.001$ for the linear effect and $p = 0.24$ for the nonlinear effect). Panel (e) shows what would happen in case the time-constant covariate total dose is considered. This last panel differs from panel (f) with respect to the time-dependent covariate, but the corresponding test of hypothesis indicates a linear functional form for total dose. Note that a regression model with total dose is used here only in order to investigate differences in how the methods are handling time-dependent and time-constant covariates. In our application about breast cancer, such a model would make no sense, as total dose can not be observed at the time origin, but only at the end of the treatment.

The proportionality of hazards was investigated by graphics and tests of hypothesis based on Schoenfeld residuals (Grambsch and Therneau, 1994). Results are discussed only for the model for the mortality rate. Violation of this assumption happens when the regression coefficients are not constant in time, as in the extended Cox model (Section 1.3.5)

$$\lambda(t) = Y(t)\alpha_0(t)\exp\left\{X^T(t)\beta(t)\right\}. \tag{2.13}$$

The presence of a time-dependent covariate in the model does not introduce any problem in the application, and results of the approach based on Schoenfeld residuals can easily be interpreted.

The assumption of proportionality was tested on the Cox model obtained at the last step of the data analysis and Figure 2.6 refers to the covariates in this model. Figure 2.6 shows, for each covariate j, a plot of the quantities $s_{jk} + \hat{\beta}_j$ against time. s_{jk} is the jth element of the scaled Schoenfeld residual at a specific failure time t_k given in equation (2.12), and $\hat{\beta}_j$ is the estimated coefficient from a standard Cox model. The plotted values $s_{jk} + \hat{\beta}_j$ are estimating the time-varying coefficient (Scheike, 2004) $\beta_j(t)$. If the proportionality assumption holds for the covariate j, i.e. its regression coefficient is time-constant, then values $s_{jk} + \hat{\beta}_j$ should be randomly distributed around a horizontal line in time. Scatterplot smooth functions in Figure 2.6 help displaying the behaviour of points in time. The assumption of a time-varying coefficient, as $\beta_j(t) = \beta_j + \theta_j t$, was verified by the test of hypothesis $H_0 : \theta_j = 0$, for each covariate j (Therneau and

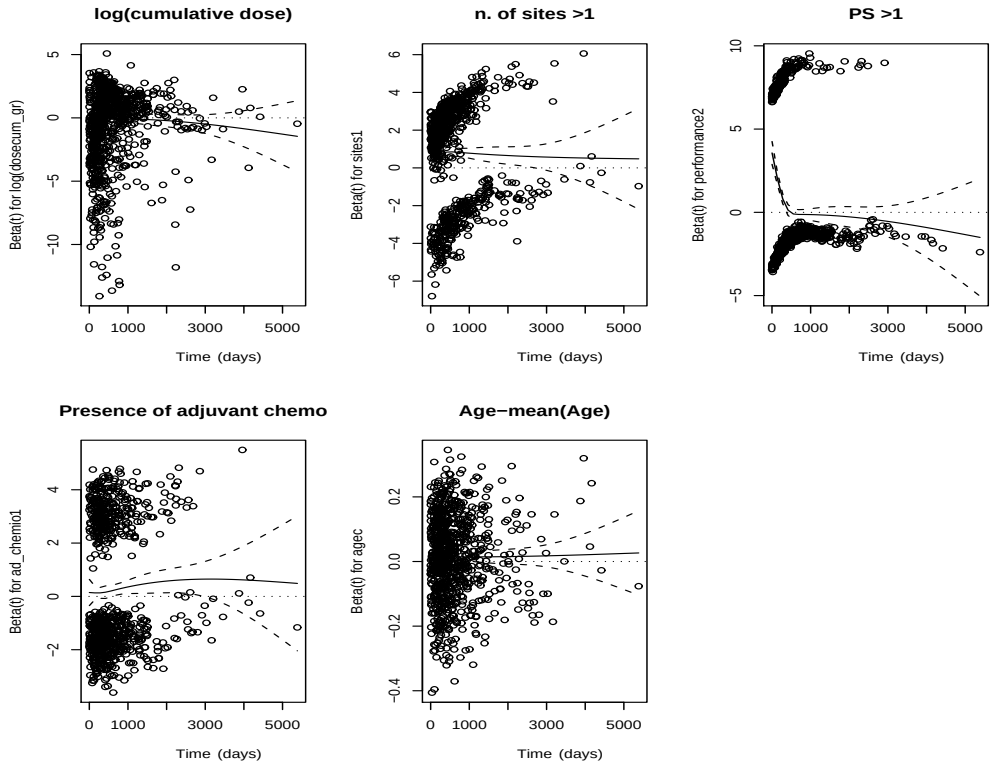

Figure 2.6: *Time-varying regression coefficients based on Schoenfeld residuals. Plots are testing for proportionality of the hazards in the Cox model for mortality.*

Grambsch, 2000, Chap. 6). Individual tests for log cumulative dose ($p < 0.001$) and PS ($p < 0.001$), as also the global test ($p < 0.001$), indicate a strong departure from proportionality of the hazards. That was the reason why the covariate PS was used as a stratification factor and a piecewise constant coefficient was assumed for cumulative dose in the final model for mortality rate.

Results about goodness-of-fit of the regression models in the application to breast cancer underline difficulties in detecting a good model and in the interpretation of graphical tests in the presence of a time-dependent covariate. This problems can be overcome by checking the model with cumulative residuals, which are various functionals of the martingale residuals (Lin et al., 1993, Wei, 1984). An introduction to these residuals is given in Section 2.1.2. If the interest is on checking the functional form of a covariate, the cumulative residuals of interest are obtained as partial sums of martingale residuals over the covariate values. For simplicity, we consider only a covariate X_1 and we distinguish between the cases of a time-dependent and time-constant covariate. In the

former case, the residual process in x is

$$M_c^1(x) = \sum_{i=1}^{n} \int_0^{\infty} I(X_{i1}(t) \leq x) \mathrm{d}\hat{M}_i(t),$$

where I is the indicator function. None of the problems previously encountered with martingale residuals, are here observed, as $M_c(x)$ is obtained as a sum over all the observations, and also over all the subjects. Each term in the process $M_c(x)$ at a certain covariate value x represents the estimated martingale $\mathrm{d}\hat{M}_i(t_x)$ of individual i on the interval $(0, t_x]$ where her time-dependent covariate $X_{i1}(t)$ assumes values smaller than or equal to x. In case the covariate is constant in time, the cumulative residuals process reduces to

$$M_c^1(x) = \sum_{i=1}^{n} I(X_{i1} \leq x) \hat{M}_i.$$

2.4 Competing risks analysis

At this step, estimates from the two Cox regression models were used for the competing risks analysis. The attention is concentrated on the cumulative incidence probability for cardiotoxicity, P_{0c}, in a a well-defined follow-up time. In the current application the time interval was chosen to be $(s, t]$. In this case the transition probability $P_{0c}(s, t)$ represents the conditional probability of developing CHF over the interval $(s, t]$, given that a patient is still alive and without cardiotoxicity at time s. The alternative choice of the entire time interval $(0, t]$ would have led to the non-conditional probability $P_{0c}(0, t)$ of developing CHF over the interval $(0, t]$. This last case would implicate different assumptions and some computational difficulties in handling the time-dependent covariate cumulative dose, as will be explained later on.

Prediction of the cumulative incidence probability requires time s and values of the covariates to be specified, i.e., a specific patient with given values of the relevant prognostic factors and of the total cumulative dose needs to be assumed. Therefore, we have fixed the covariates of both the Cox regression models and called them $X_c^{(0)}$ and $X_d^{(0)}$. Consequently, using the formula in (2.3) we have computed the estimators of the two cumulative cause-specific hazard increments as follows:

$$d\hat{A}_h(t; X_h^{(0)}) = d\hat{A}_{h0}(t, \hat{\beta}_h) \exp\left\{(X_h^{(0)})^T \hat{\beta}_h\right\}, \quad h = c, d. \tag{2.14}$$

Vectors $\hat{\beta}_c$ and $\hat{\beta}_d$ contain parameter estimates in the Cox models for the cardiotoxicity

hazard and the mortality hazard, respectively.

In the application we have fixed time s equal to six months and we have assumed that s represents the end of treatment. Therefore, cumulative dose, which is administrated only during the treatment period, is time-independent in $(s, t]$ and equal to the total dose at the end of treatment. Formally we are in a situation where $X_1(u) = X_1(s)$ for all u such that $s \leq u \leq t$. That is why the covariate vectors $X_h^{(0)}$, $h = c, d$, which include cumulative dose X_1, are constant in time and easily fixed in (2.14) for computing $d\hat{A}_h(t; X_h^{(0)})$.

Through the plug-in method, increments of the estimated cumulative hazards for cardiotoxicity in (2.14) and the estimated survival function were used to compute the Aalen-Johansen type estimator in (2.4), as follows:

$$\hat{P}_{0c}(s, t; X_c^{(0)}, X_d^{(0)}) = \sum_{s<t_{cj}\leq t} \hat{P}_{00}\left(s, t_{cj}-; X_c^{(0)}, X_d^{(0)}\right) \mathrm{d}\hat{A}_c(t_{cj}; X_c^{(0)}). \quad (2.15)$$

We underline that this estimator works correctly in case all the covariates are time-independent in the considered time interval $(s, t]$.

In our application we computed $\hat{P}_{0c}(s, t)$ as a function of dose and time, with $s = 0.5$ and $t = 2.5$ years. Predictions on CHF probability were made for different levels of total epirubicin (600, 800, 900 and 1000 mg/m^2) and for some different combinations of values of the prognostic factors. As an example, risk of developing CHF from 0.5 to 2.5 years of follow-up for a patient without risk factors, with performance status equal to 1 and number of tumour sites higher than 1 is shown in Figure 2.7 for age 40, 50, 60 and 70. Each risk curve on $(s, t]$ is associated with a fixed level of cumulative dose.

For all possible typologies of patients, probability of developing CHF increased mostly during the first eight months after stopping treatment, becoming nearly constant at the end of the 2.5 years follow-up, as it is also shown by the example in Figure 2.7. The cardiotoxicity risk increased with age for fixed doses. Moreover, the substantial increase in the risk of developing CHF, as the cumulative dose rose from 600 mg/m^2 up to 1000 mg/m^2, was highest for older patients, as the risk curves became gradually more spaced from age 40 to age 70 (Figure 2.7).

In this application, cumulative incidence probabilities were rich of information from a medical point of view, because they yield risk of cardiotoxicity at each fixed time point u, for $u \in (s, t]$, and for each combination of the significant risk factors in the competing risks model. Table 2.3 shows a numerical example of increasing risk for older patients and for higher total doses, in case of presence of previous antihormonal

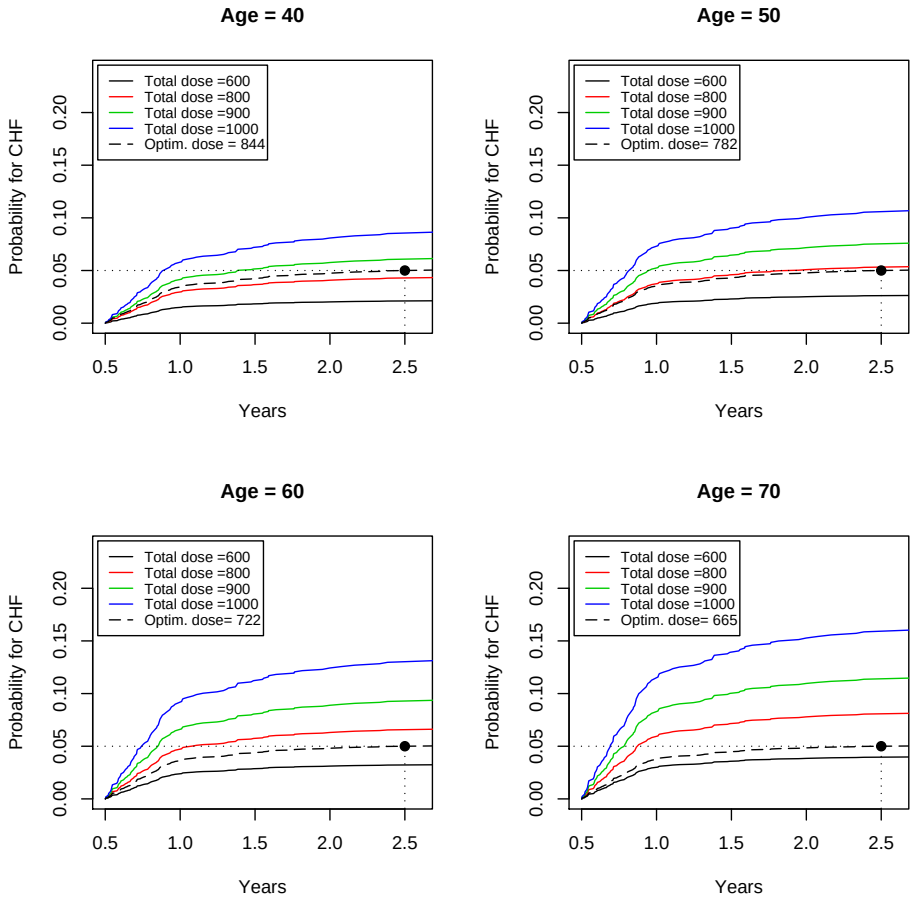

Figure 2.7: *Risk of developing CHF from 0.5 to 2.5 years of follow-up at age 40, 50, 60 and 70 for patients without risk factors, with number of sites >1 and performance status=1. The solid black, red, green and blue lines represent the probability curve associated with an epirubicin treatment of 600, 800, 900 and 1000 mg/m^2, respectively. The dashed black line represents the curve of CHF risk corresponding to the optimal recommended cumulative dose; the mark along that curve indicates the 5% probability level reached at 2.5 years.*

treatment, performance status $= 1$ and number of tumour sites $= 1$.

2.5 The optimal recommended dosage at 5% risk for cardiotoxicity

In the medical literature the recommended epirubicin total dose is based on a 5% estimated risk of CHF (Ryberg et al., 1998). Therefore, the maximal level of total dose (mg/m^2) associated with an estimated probability of developing CHF equal to 5% was computed for each combination of values of prognostic factors. A total dose below

Risk of cardiotoxicity at time t					
Time	**Cumulative dose (mg/m^2)**	**Age**			
		40	*50*	*60*	*70*
$t = 547.87$ (1.5 years)	*600*	0.04	0.05	0.06	0.07
	800	0.07	0.09	0.12	0.14
	900	0.10	0.13	0.16	0.20
	1000	0.14	0.17	0.22	0.26
$t = 913.12$ (2.5 years)	*600*	0.05	0.06	0.07	0.09
	800	0.09	0.11	0.14	0.17
	900	0.13	0.16	0.19	0.23
	1000	0.17	0.21	0.26	0.31

Table 2.3: *Risk of cardiotoxicity as a function of cumulative dose (600, 800, 900 and 1000mg/m^2) at specific time points $t = 1.5$ and $t = 2.5$ (years) in case of tumour sites $= 1$, presence of previous antihormonal treatment, performance status $= 1$, for age 40, 50, 60, and 70..*

that maximal predicted threshold would assure a risk lower that 5% to develop cardiotoxicity. We have denominated the predicted threshold of maximal total dose as the optimal dosage.

In order to find optimal dosages for each typology of patient, from (2.1) we consider the cumulative incidence probability

$$P_{0c}(s, t_0; X_1) = \int_s^{t_0} S(u-; X_1)\alpha_c(u; X_1)du,$$

where $S(t; X_1)$ is the survival function in (2.2), as a function of the total cumulative dose, X_1, on which both the cause-specific hazards depend. We fixed a certain time $t = t_0$. The variable X_1 is a component of both the covariate vectors X_c and X_d. The predicted threshold for the total dose is computed by holding fixed all the covariates at $X_c^{(0)}$ and $X_d^{(0)}$, except the cumulative dose X_1. In this case, we consider the estimate $\hat{P}_{0c}(s, t_0; X_1, X_c^{(0)}, X_d^{(0)})$ in (2.15) only as a function of the total cumulative dose X_1. The optimal recommended dose is then the numerical solution $X_1 = X_1^*$ to

$$\hat{P}_{0c}(s, t_0; X_1^*, X_c^{(0)}, X_d^{(0)}) = 0.05.$$

Note that if we fix the covariates at values other than $X_c^{(0)}$ and $X_d^{(0)}$, we obtain a different recommended total dose.

	Case	Risk factors	Age			
			40	50	60	70
Performance status $= 1$	A:	No risk factors	806/844	739/782	673/722	609/665
	B:	CMF for relapse	864/883	828/850	793/818	759/786
	C:	Previous Tam	626/670	561/610	496/552	434/496
	D:	Irradiation Spine/med	596/640	530/581	467/523	404/467
	E:	Disposition Heart Disease	491/539	427/481	364/424	303/369
Performance status > 1	F:	No risk factors	890	835	783	732
	G:	CMF for relapse	908	878	848	820
	H:	Previous Tam	723	670	620	571
	I:	Adjuvant CMF	917	865	815	767
	J:	CMF for relapse + adjuvant CMF	922	893	866	839

Table 2.4: *Optimal recommended total dosages by performance status (PS) and age. Doses about patient typologies with PS* $= 1$ *are given for number of sites* $= 1/ > 1$. *Doses about patient typologies with PS* > 1 *are given only for number of sites* > 1.

2.5.1 Examples

In order to show how results about the cumulative incidence probability and the optimal dosages were obtained, we consider the following example. We choose a typical patient, for instance, a patient 50 years old with number of tumour sites > 1, a performance status equal to 1, with disposition to heart disease and antihormonal treatment as the only risk factors. Formally, it corresponds to choosing

$$X_2 = X_3 = 1, \qquad X_4 = X_6 = X_7 = 0, \qquad X_5 = 50 \text{ years},$$

$$X_8 = 1 \text{ (n. of sites} > 1), \qquad X_9 = 1.$$

We remark that performance status is a stratification factor and it does not have any regression coefficient. It affects the estimate of the baseline cumulative hazard for mortality, and therefore, it must also be fixed when estimating the cumulative incidence probability. If we also fix a value for the total cumulative dose, for instance, 800 mg/m^2, the corresponding estimate of the probability is $P_{0c}(s,t) = 0.26$. For other choices of total dose, 600, 900 or 1000 mg/m^2, the estimated probability is, respectively, 0.14, 0.35 and 0.45. On the other hand, if we intend to find the optimal total dose for this type of patient, we fix a time interval $(s, t_0]$ with $t_0 = 2.5$ years. Then, we compute the value of the total dose which makes the estimated probability $\hat{P}_{0c}(s,t_0)$ equal to 5%. In the example, the result is 312 mg/m^2.

Table 2.4 shows some results about the optimal dosage recommended in an epirubicin treatment of six month ($s = 0.5$ years) for each typology of patient. Results in the table are classified by performance status and number of sites ($= 1$ or > 1). The time

interval $(s, t_0]$ was fixed equal to $(0.5, 2.5]$ years. For patients without risk factors, with number of sites $= 1$ and performance status $= 1$, the cumulative dose which assured a 5% risk of developing CHF was equal to 806 mg/m^2 at age 40 and decreased gradually with increasing age, being equal to 609 mg/m^2 at age 70 (patient A in Table 2.4). If number of sites is > 1, the cumulative dose increased slightly, being 844 at the age 40 and 665 at age 70. Curves of cardiotoxicity risk are shown in Figure 2.7 for these last patients with number of sites > 1.

Probability of developing CHF depends strongly on the risk factors for this disease. That is why results about the optimal dose can vary greatly according to which risk factor is present. As the oldest patients had the highest substantial increase in CHF probability, their optimal dosage is about 200 mg/m^2 less compared to the youngest patients for almost all the patient typologies. Cases with CMF for relapse are an exception (patient B in Table 2.4). Some risk factors are found to be more severe for risk of cardiotoxicity than others and lower doses are then recommended in their presence. This observation can be noted from Table 2.4, where, in case of performance status $= 1$, presence of previous antihormonal treatment (patient C) reduces the recommended dosage compared with the case of no risk factors (patient A). Moreover, presence of irradiation to the spine lowers the optimal dose further. Finally, presence of disposition to heart disease appears to be the most severe risk factor, as it corresponds to the lowest suggested doses.

We expect patients with some risk factors for cardiotoxicity to have an optimal dosage lower than the one for patients without any risk factor. This idea did not hold in case of CMF for relapse (patient B in Table 2.4), because its effect needs to be interpreted taking into account the interaction with the cumulative dose, i.e. in combination with a specific cumulative dose. Presence of CMF reduces the cardiotoxicity rate in combination with doses lower than 928 mg/m^2 (Figure 2.3) and the optimal dose resulted to be lower than 900 mg/m^2 in all the cases. Therefore, patients with CMF for relapse are associated with high optimal doses, even higher than the ones in the case of no risk factors, as shown in Table 2.4. A total dose equal to 864/883 mg/m^2 is associated with 40 years and a decrease of only about 100 mg/m^2 is observed for the oldest patients with 70 years (patient B), in contrast with the cases mentioned above.

The consequences of including the competing risk of dying for breast cancer into the study can be noted with various examples. Differences in the optimal dosages between a low performance status ($= 1$) and a more severe status (> 1) are due to the competing cause of death, because for patients with status > 1, who have a higher risk of dying,

the event of cardiotoxicity is less probable to be observed. For example, a patient with very high mortality rate from breast cancer, as the case I in Table 2.4, is associated with a higher optimal dose than the recommended dose for a patient with lower risk of dying, as the case A. The same mechanism about the competing risks is observed when comparing patients with number of sites $= 1$ and > 1. From a medical point of view, the recommended doses need to be increased in order to respond to the more severe status of patients who have numerous tumour sites.

Comparison between the case E and the case I in Table 2.4 shows the importance of taking the risk factors for both causes into account and how they can be balanced. In fact, the difference in the dose, which is more than 400 mg/m^2, is attributable to the high cardiotoxicity rate for patient E and the increased mortality rate of patient I.

2.6 The time-dependent cumulative dose and its interpretation

In this section we recall some assumptions used in the application to breast cancer in predicting the cumulative incidence probability for cardiotoxicity. Moreover we discuss how the time-dependent covariate cumulative dose was defined and handled, and whether this covariate leaves the related inference unchanged. Finally, we discuss the prediction of the cumulative incidence probability when some alternative assumption are made.

In the application to breast cancer, the covariate cumulative dose was named as $X_{c,1}(t)$ and $X_{d,1}(t)$ in the Cox models for the cardiotoxicity rate and the mortality rate, respectively. Since both $X_{c,1}(t)$ and $X_{d,1}(t)$ represent the same covariate, but as elements of the different vectors X_c and X_d, here we will simplify the notation by calling them $X_1(t)$.

In Section 2.4 we estimated the cumulative incidence probability for cardiotocixity as follows. Since the chemotherapy treatment period was assumed to be equal to $(0, s]$, with s being a fixed time within the observation period $[0, \tau]$ of the study, and since the dose was administrated only along the treatment period, the covariate $X_1(t)$ is time-dependent only within the time interval $(0, s]$, whereas it is time-independent and equal to the total dose $X_1(s)$ in the remaining period $(s, \tau]$. Since an interesting medical aspect consists of studying the risk of cardiotoxicity after ending the chemotherapy treatment (after time s), we decided to predict the behaviour of the cumulative inci-

dence probability between the given time s and a time t, with $t > s$. Therefore, the estimate of this probability by equation (2.15) was straightforward because just a single given value for the cumulative dose X_1, being time-independent in $(s, t]$, was required. Implementation of equation (2.15) did not imply any computational difficulty.

If the interest was on studying the risk of cardiotoxicity from the beginning of the treatment until a given time t, $t > s$, then X_1 would no longer be time-independent on the whole interval of interest $(0, t]$, and the specification of a single covariate value would not be sufficient anymore. This alternative scenario shall be illustrated later in this section, after formally defining the time-dependent covariate structure.

Let the process $\{X_1(t), 0 \leq t \leq \tau\}$, where $[0, \tau]$ is the observational period of the study, be the time-dependent covariate cumulative dose. In the application to breast cancer it turns out that the process $X_1(t)$ can be condidered as deterministic, i.e. fixed in advance. This means that the covariate cumulative dose belongs to the class of so-called external defined time-dependent covariates (Kalbfleisch and Prentice, 2002, Chap. 6). Therefore, the form of the likelihood function and inference are unchanged, and predictions of the cumulative incidence probabilities in the competing risks regression model can be performed without any complication. The reason is that the administration of the cumulative dose along the treatment period is assumed to be regulated, for each patient, by a predetermined time schedule between the possibilities shown in Table 2.1.

Unlike the competing risks regression model of our application, generally, in regression models covariates are considered as random variables. The usual assumption is that the hazards functions refer to the conditional distribution of survival times given the observed covariates. Therefore, it appears to be natural to consider a time-dependent covariate as a stochastic process. Nevertheless, handling such a process in competing risks regression models is not always straightforward, and in some cases it is even not possible to make predictions (Fisher and Lin, 1999, Andersen et al., 1993, Chap. 3). Further explanations are given in Appendix B of the thesis.

After having specified the type of process $X_1(t)$ for the cumulative dose, we return to the hypothetical situation of studying the risk of cardiotoxicity over the entire time interval $(0, t]$. We shall briefly discuss how the predictions of the risk as a function of the cumulative dose are obtained. In order to be able to predict the cumulative incidence probability for cardiotoxicity (cause $h = c$) by equation (2.6), given the data and the estimates $\hat{\beta}_h$ of the regression coefficients, we need to specify a path of the process $X_1(t)$, besides the values for the remaining covariates in the vectors X_c and

X_d. Denote the given path for the cumulative dose by $X_1^{(0)}(t)$. This path consists of a non-decreasing left-continuous step function with jump size equal to the single dose injected; its maximum value is reached at the end of the treatnment (time s) and hence it is equal to the total dose $X_1^{(0)}(t)$.

In estimating the cumulative imcidence probability for cardiotoxicity, $P_{0c}(0,t)$, in presence of the cumulative dose process $X_1(t)$, the theory illustrated in Section 2.1.1 and Section 2.4 is still valid and the formulas need just to be updated by substituing X_h with $X_h(t)$. Equation (2.4) expresses still valid estimators for the cumulative cause-specific baseline hazards, the only differences being in the evaluation of $S_h^{(0)}(t,\hat{\beta}_h)$. The latter formula needs also the values of the cumulative dose for all patients at risk at that time t and therefore, this fact might yield some computational difficulties in the estimation of the survival probability in (2.5). Moreover, the values of the given path $X_1^{(0)}(t)$ at all the cause-specific failure times are needed in order to compute Equations (2.14) and (2.15).

Optimal recommended dosages might be also investigated when the interest focuses on the risk of cardiotoxicity from the beginning of the treatment until a given time t after the treatment. Let us illustrate a possible procedure by an example. Similarly to what was done in Section 2.5, we may consider the cumulative incidence probability $P_{0c}(0,t_0)$, for a given time t_0, as a function of the cumulative dose, i.e., as a function of the deterministic process $X_1(t)$. Furthermore, we might restrict the attention to the case of such a functional relationship under the assumption that a certain time schedule is specified. For instance, we assume a time schedule of an injection at day one every three weeks (Table 2.1) during a treatment period $(0,s]$, with $s < t_0$. An interesting aim would then be to find the single dose injected every three weeks, which assures a cardiotoxicity risk lower than 5% at time t. In order to solve this problem and find such a recommended single dose, it would be sufficient to find the corresponding recommended total dose as a numerical solution to the equation $\hat{P}_{0c}(0,t_0) = 0.05$. Nevertheless, the optimal total dose found by this procedure, or, equivalently, the corresponding optimal single doses, would be strictly related to the assumed time schedule and, therefore, could not be interpreted by its own. In conclusion, this hypothetical analysis would compute a dose administration regime which is optimal for a certain given time schedule (three weeks in the previous example) and with respect to a 5% treshold for the cardiotoxicity risk. Further recommended dose administration regimes might also be determined for the different time schedules in the study, in order to provide a general picture and useful medical guidelines about the relation between the

chemotherapy treatment and cardiotoxicity.

2.7 Discussion

The competing risks setting was chosen as a very necessary statistical tool for studying the cardiotoxicity risk for patients with advanced breast cancer. Because of their severe status, it is known that these patients have a very high risk of dying, also during their chemotherapy treatment. That is why we can not neglect to consider the competing risk of dying for breast cancer even though the primary interest focuses on risk of developing CHF. Ignoring the competing cause might lead to overlooking important features of the studied problem. Patients who died could potentially have developed CHF, but this event can never be observed. The comments on the numerical results about the optimal total doses described in Section 2.5 investigate the mechanism underlying the two competing causes and describe a possible reasonable interpretation of the medical problem.

The application of a competing risks analysis to the study of cardiotoxicity as a function of chemotherapy dosages led to very important new medical results. First of all, we found new recommended levels for the total dose administrated during Epirubicin chemotherapy, which were found to be lower than the one recommended in the literature (Ryberg et al., 1998). Moreover, the existing literature suggests a single level for all types of patients. We demonstrated that the optimal recommended dosage can vary substantially between groups of patients with different characteristics and risk factors.

In order to compute the optimal dosage levels corresponding to a 5% cardiotoxicity risk, we needed to treat the cumulative dose as a time-dependent covariate. Handling the time-dependent cumulative dose turned out to be easy as it was considered as a deterministic process. However, as the history of dose administration for each patient was needed, the implementation of the analysis was not trivial. In general, handling time-dependent covariates required particular attention since it is essential to define which kind of process is underlying the covariate (Kalbfleisch and Prentice, 2002, Chap. 6).

The standard method for competing risks regression models (Andersen et al., 2002, 1993, Chap. 7) was used for the statistical analyses. Although new alternative methods (Scheike and Zhang, 2004, Fine, 2001, Scheike et al., 2007, Andersen et al., 2003) have appeared recently in the literature, the standard approach enabled us to perform a complete and accurate analysis, as selection and goodness-of-fit of the model follow

methods which are well-established in survival analysis and applicable to this context. Moreover, standard software for survival analysis can be used for regression models for the cause-specific hazards by regarding all events due to other causes than the one of interest as additional censoring events. Nevertheless, this idea about censoring is correct only when analyzing cause-specific hazard functions and cumulative incidence probabilities, while it yields erroneous conclusions if it is used in computing Kaplan-Meier type estimates for the single causes. These estimates would not be equal to one minus the cause-specific estimated cumulative incidence probabilities (Tsiatis, 1975, 1998).

A drawback of the standard method for regression analysis of competing risks data is that simple parameters, which explain directly the effects of covariates on the cause-specific cumulative incidence probabilities, are missing. The cumulative incidence probabilities are complex non-linear functions of the covariates and therefore it is only possible to describe indirect covariate effects by estimating these probabilities for different given covariate patterns.

In our breast cancer study, the interest was limited to estimating the cumulative incidence probabilities in the interval $(s, t]$ with $s = 0.5$ and $t = 2.5$ years. An interesting suggestion might be to investigate and compare the cardiotoxicity risk in different time intervals in order to identify periods with highest risk. This can be performed by considering the conditional probabilities $P(T \leq s + \Delta, Z(T) = c \,|T \geq s)$, for instance with $\Delta = 0.5$ years and $s = 0.5, 1, 1.5, 2$ years. For the notation of these conditional probabilities the reader can refer to equation (1.39).

Problems about goodness-of-fit in case of a time-dependent covariate were investigated. Some of them were already pointed out by other authors (Therneau and Grambsch, 2000, Chap. 5), but we disagree on the usefulness of martingale residuals in suggesting possible correct functional form. Plots of martingale residuals both per-observation and per-subject might fail in investigating the functional form of a time-dependent covariate. We discussed about the need of cumulative martingale residuals (Lin et al., 1993) in model diagnostics, as they overcome problems related to time-dependency of covariates. A drawback of the type of residuals discussed in this chapter and the corresponding tests of hypotheses for each covariate, is that they are only valid if the Cox model is correct for all the remaining covariates (Scheike and Martinussen, 2004).

Chapter 3

Time-varying Regression Coefficients in Relative Survival Models

In relative survival modelling through regression analysis, the existing approaches can be classified within the parametric, semiparametric and nonparametric settings. Here we present the additive excess hazards models (Zahl, 1996), where the excess hazard is on additive form. We assess the importance of time-varying effects for regression models in this framework and show how recent developments can be used to make inferential statements within the nonparametric version of the model. When some covariate effects are constant, we show how the semiparametric additive risk model can be considered in the excess risk setting, providing a better and more useful summary of the data. Estimators having an explicit form and inference based on a resampling scheme are presented for both the nonparametric and semiparametric models. We also describe a suggestion for goodness-of-fit of relative survival models, which consists of statistical and graphical tests based on cumulative martingale residuals. This is illustrated on the semiparametric model with proportional excess hazards. We analyze data from the TRACE study using different approaches and show the need for more flexible models in relative survival.

The research work presented in this chapter is based on the paper Cortese and Scheike (2008).

3.1 Introduction and background

3.1.1 Relative survival

In many cancer studies, but also in population-based and clinical observational studies other than cancer, information on causes of death, remissions, etc. is sometimes unavailable, especially with a long follow-up. In some cases, this information is recorded on medical registries but it is incomplete or misleading, because death could be only partially due to the disease of interest and it is difficult to classify deaths due to other causes indirectly correlated with the disease of interest. For this reason, the use of cause-specific survival in the framework of competing risks, where at least two distinct alternative causes need to be specified, is problematic. Moreover, many clinical studies aim at identifying prognostic factors for mortality due to the disease, differentiating whether their effects are also related to the natural mortality in the underlying population. In this case, problems arise in comparisons between studies based on different background populations.

Relative survival analysis provides a solution to these difficulties. It does not require information on cause of death, whereas it allows one to estimate patient survival corrected for the effect of other causes of death, using the natural mortality of the underlying population. Of course, the natural mortality encompasses also mortality from the disease of interest; however, when the latter is very small and then negligible, the general population is commonly assumed to be unaffected by the disease of interest. Indeed, relative survival describes the excess mortality for patients diagnosed with the disease of interest, irrespective of whether the excess mortality is directly or indirectly attributable to the disease. In general, estimation of this corrected patient survival, a quantity which is hypothetically defined as the net survival in the competing risks setting, is the principal aim in relative survival. From population life tables theory, the estimate is given by the relative survival ratio between the observed survival of patients and the expected survival from the underlying population, with respect to the main factors affecting the natural mortality, such as age, sex and calendar time.

A natural way of modelling relative survival through regression analysis consists in assuming the following additive form for the hazard at time t, conditional on covariates Z and X:

$$\lambda(t; Z, X) = \lambda^*(t; Z) + \nu(t; X), \tag{3.1}$$

where Z and X are sets of covariates which are not necessarily all distinct. The total

observed hazard λ is modelled as the sum of the expected hazard $\lambda^*(t; Z)$, which represents the background rate of mortality of the general population, and the excess hazard $\nu(t; X)$ due to presence of an additional cause of mortality, such as cancer or other chronic diseases. The expected hazard is generally estimated from external data, i.e., mortality rates recorded in the public registries of the population underlying the patients' sample under study. It is assumed to be known in the relative survival model and generally it depends on some characteristics Z of the population. The additional excess hazard follows a regression model based on the relevant risk factors X and can be modelled by a proportional or an additive form, according to the validity of the underlying assumptions. In general, the principal interest in regression analysis consists in evaluating possible prognostic factors which influence directly the excess risk, in absence of the effect of competing causes of death. That is why only the excess risk is supposed to depend on the set of covariates observed in the exposed individuals.

3.1.2 Parametric, semiparametric and nonparametric approaches

Among different approaches to modelling relative survival, our attention is directed to models following the additive form in (3.1). Within this approach, various models and their extensions have been proposed recently and they can be classified as parametric, semiparametric or nonparametric models. Two basic methods that assume a multiplicative function of the covariates for the excess hazard, described by Hakulinen and Tenkanen (1987) and Estève et al. (1990), have been used in the parametric setting. Extensions of these models (Dickman et al., 2004, Lambert et al., 2005) and handling time-dependent covariates (Bolard et al., 2001) have also been developed in the literature. Although all these models are specified in continuous time, they assume a parametric function for the hazard, usually a constant hazard within predetermined time-intervals. In order to detect possible nonproportional excess hazards, the standard solution used within these models consists of including time-dependent covariates as interaction terms (covariate by follow-up time-intervals). More recently, some suggestions have used spline functions (Giorgi et al., 2003, Bolard et al., 2002) for modelling time-dependent hazard ratio and the baseline excess hazard, in order to yield more flexible and less restrictive additive models, in case of multiplicative scale for the excess hazard. In the semiparametric setting, these attempts can be seen as alternatives to the well-known proportional excess hazards model by Sasieni (1996). The semiparametric proportional excess hazards model considers an excess risk on Cox form and can easily handle time-dependent covariates, provided that the assumption of a propor-

tional hazards for the excess risk of individuals is verified. Zahl (1996) considered the fully nonparametric additive hazards model (Aalen, 1980) described in Example 1.3.1, to model the excess hazard, with $\lambda(t; X) = \alpha_0(t) + \beta_1(t)X_1 + \ldots + \beta_p(t)X_p$, overcoming problems about non-proportionality and non-positive excess hazards (Zahl and Tretli, 1997, Zahl, 1995).

3.1.3 Dynamic extensions for the nonparametric and semiparametric settings

We shall study the additive hazard models and show how recent developments can be used to make inferential statements within the nonparametric additive excess hazards model. This makes it possible to test the key hypothesis that an excess risk effect is time-varying in contrast to being constant over time. One problem with the fully nonparametric dynamic description is that the model might be too big, if some covariate effects are in fact constant with time. We shall therefore also show how the semiparametric additive risk model (McKeague and Sasieni, 1994) can be considered in the excess risk setting. This model can provide a better and more useful summary of the data and makes a better bias/variance trade-off. We shall show how these two additive models are easy to fit with estimators on explicit form and how inference including tests for time-constant effects can be carried out based on a resampling scheme.

Our objective is to introduce and to assess the importance of time-varying effects (Scheike, 2004) for regression models in the relative survival framework. Their presence in the model shows directly how the influence of risk factors on the excess hazard may change over follow-up time, as regression coefficients are allowed to depend on time. No difficulties appear in handling time-dependent covariates, which are treated as commonly performed in the Aalen additive hazards model and in the Cox model.

3.2 The nonparametric additive excess hazards model

The nonparametric additive excess hazards model, described by Zahl (1996), is

$$\lambda(t) = Y(t)[\alpha^*(t; Z) + X^T(t)\beta(t)], \tag{3.2}$$

and contains only nonparametric terms. The excess rate is modelled in additive form and follows the additive hazards model (Example 1.3.1) introduced by Aalen (1980). The p-dimensional vector of covariates is denoted by $X(t)$. The function $Y(t)$ is

the risk indicator, which is one if the event or censoring has not occurred until t and zero otherwise. The effects of risk factors on the excess mortality hazard $\nu(t; X) = Y(t)X^T(t)\beta(t)$ are expressed by the time-varying p-dimensional regression coefficient $\beta(t) = (\beta_1(t), \ldots, \beta_p(t))^T$.
The relative survival for the additive excess hazards model is equal to

$$r(t) = \exp\left\{-\int_0^t X^T(s)\beta(s)ds\right\},$$

which in general, for additive models, can be written as $r(t) = S(t)/S^*(t)$, where $S(t)$ and $S^*(t)$ denote the observed and expected survival, respectively.

3.2.1 Notation

The models and the related inference are given using the counting process representation described in Section 1.1.3. The conditional intensity $\lambda(t)$ in (3.2) provides a model for its associated counting process $N(t)$, that counts the observed failures in the observation period $t \in [0, \tau]$, with $\tau < \infty$, of a subject with predictable covariates Z and X. Let $(N_i(t), Y_i(t), Z_i(t), X_i(t))$ for $i = 1, \ldots, n$, be n independent observations from the additive excess hazards model with intensity $\lambda(t)$. Recalling the definition (1.17) in Chapter 1, denote by $N(t) = (N_1(t), \ldots, N_n(t))^T$ the multivariate counting process of the n subjects, and by $\lambda(t) = (\lambda_1(t), \ldots, \lambda_n(t))^T$ the associated intensity. The $n \times p$ dimensional matrix $\boldsymbol{X}(t) = (Y_1(t)X_1(t), \ldots, Y_n(t)X_n(t))^T$ contains all the information about the predictable covariates in the excess rate. The considered estimators have properties that rely on the martingale theory described in Section 1.2.1. $M(t) = N(t) - \Lambda(t)$ is the n-dimensional zero-mean martingale associated with the counting processes $N(t)$. The total cumulative intensity is given by the compensator of the martingale, $\Lambda(t) = \int_0^t \lambda(s)ds$. Define $\lambda^*(t) = (Y_1(t)\alpha_1^*(t), \ldots, Y_n(t)\alpha_n^*(t))^T$. We then have the increments $dN(t) = \lambda(t)dt + dM(t) = \lambda^*(t)dt + \boldsymbol{X}(t)\beta(t)dt + dM(t)$ of the counting process using the excess additive regression model.

3.2.2 The estimators

Inference is made by estimating the cumulative regression coefficients $B(t) = \int_0^t \beta(s)ds$, which give the cumulative effects of each covariate on the excess mortality rate. Estimators in the additive excess hazards model are very similar to the ones used for the standard additive hazards model (Example 1.3.1). Let $\Lambda^*(s) = \int_0^t \lambda^*(s)ds$. The prin-

cipal basic difference in working with relative survival consists in replacing the usual counting process $N(t)$ with the modified counting process $\tilde{N}(t) = N(t) - \Lambda^*(t)$. Thus from the increments of the martingale, we have

$$d\tilde{N}(t) = dN(t) - \lambda^*(t)dt = \boldsymbol{X}(t)\beta(t)dt + dM(t),$$

which suggests the possibility to estimate the increments $\beta(t)dt$ by weighted least squares methods for multiple linear regression (Huffer and McKeague, 1991). The increment in $\tilde{N}_i(t)$ thus gives the observed excess risk compared with the background mortality, among those under risk, since the martingale increment has mean zero. In other words, the expected number of deaths equals the expected number of background deaths plus the expected number of excess mortality deaths. The resulting estimator is

$$d\hat{B}^*(t) = \boldsymbol{X}^-(t)d\tilde{N}(t), \tag{3.3}$$

where the $p \times n$ matrix

$$\boldsymbol{X}^-(t) = (\boldsymbol{X}^T(t)W(t)\boldsymbol{X}(t))^{-1}\boldsymbol{X}^T(t)W(t) \tag{3.4}$$

is the generalized inverse of $\boldsymbol{X}(t)$ and $W(t)$ is a predictable $n \times n$ diagonal matrix of weights. Therefore,

$$\hat{B}^*(t) = \int_0^t \boldsymbol{X}^-(s)d\tilde{N}(s) \tag{3.5}$$

is the estimator for the p-dimensional vector of cumulative regression coefficients.

The estimator in (3.5) can be written as

$$\hat{B}^*(t) = \hat{B}(t) - \int_0^t \boldsymbol{X}^-(s)\lambda^*(s)ds, \tag{3.6}$$

the difference of the standard Aalen estimator (Aalen, 1980), $\hat{B}(t) = \int_0^t \boldsymbol{X}^-(s)dN(s)$, and a predictable term depending on the known background mortality rate and the observed covariates (Zahl and Tretli, 1997). The second term represents the average expected hazard of the population at risk at each observed time, weighted with the observed covariate values. The Aalen estimator is incremented at each failure time (where a jump is observed) while it is constant between failures. Note that the estimator $\hat{B}^*$ decreases systematically between failure times because of the Lebesgue integral in the second term of (3.6). In this latter $d\tilde{N}_i(t)$ is negative for $t < T_i$ (T_i is the failure time of the individual i) and, at failure times, it is observed to have jumps

equal to $N_i(T_i) - \Lambda_i^*(T_i)$. Moreover, the estimators depend on both censoring and failure times, as the modified counting process $\tilde{N}(t)$ for censored individuals depends on their precise censoring times (while the observed counting process $N(t)$ is constantly equal to zero for censored subjects). Consequently, even though the estimator $\hat{B}^*$ is well-defined by expression (3.3) and is an unbiased estimator of the excess mortality, some care has to be taken when implementing the Lebesgue integration. Even though the substitution of integrals with summations might require further assumptions, in practical cases where the expected hazard λ^* is piecewise constant, this substitution is allowed. This aspect is also discussed in related papers (Andersen and Væth, 1989, Zahl, 1996, Sasieni, 1996).

The approximate maximum likelihood estimator for $B(t)$ is

$$\tilde{B}^*(t) = \int_0^t \boldsymbol{X}_W^-(s) d\tilde{N}(s),$$

where $\boldsymbol{X}_W^-$ is the matrix in (3.4) with diagonal weight matrix $W(t) = \text{diag}(Y_i(t)/\lambda_i(t))$. The estimator is obtained from score equations for the infinite-dimensional parameter $\beta(t)$ as in Greenwood and Wefelmeyer (1991), Sasieni (1992). For a more general theory on estimation equations for infinite-dimensional parameters the reader is referred to Greenwood and Wefelmeyer (1990). Since the matrix $W(t)$ contains the unknown parameter $\beta(t)$ through $\lambda_i(t)$, $\tilde{B}^*(t)$ requires estimating $W(t)$, which can be preformed by the application of smoothing techniques. The estimated weight matrix is then plugged into the estimator $\tilde{B}^*(t)$. $\tilde{B}^*(t)$ is asymptoticly efficient and $\sqrt{n}$ times its difference with the true cumulative regression parameter converges in distribution to a Gaussian martingale, as for the least squares estimator $\hat{B}^*(t)$.

3.2.3 Properties of the estimators

If the matrix $\boldsymbol{X}(t)$ has full rank for all t, $\hat{B}^*(t)$ is an unbiased estimator of $B(t)$, because the second term in

$$\hat{B}^*(t) = \int_0^t dB(s) + \int_0^t \boldsymbol{X}^-(s) dM(s)$$

is a martingale with zero mean. Moreover, using functional forms of the strong law of large numbers, under certain regularity conditions the following convergence in distribution can be proved (Martinussen and Scheike, 2006, Chap. 5):

$$n^{1/2}(\hat{B}^* - B) \xrightarrow{D} U, \quad \text{for } n \to \infty,$$

where U is a Gaussian martingale with covariance function $\Phi(t) = \int_0^t \phi(s)ds$. An explicit expression for $\phi(t)$ can be found in Martinussen and Scheike (2006, Chap. 5). As a general reference within the thesis, the asymptotic theory is described in Section 1.2.3. These simple properties of the estimator $\hat{B}^*$ are the fundamental elements for inference and are the same as those for the estimator $\hat{B}$ for the standard nonparametric additive hazards model (Example 1.3.1), since the asymptotic results are still based on the martingale $M(t)$. The martingale in the additive excess model differs from the one in the standard additive model only for the expression of its compensator $\Lambda(t)$. In fact, a component of this latter is constrained to be equal to the integrated expected mortality of the population.

One of the possible estimators for the variance of $\hat{B}^*$ is

$$\hat{\Phi}(t) = n \int_0^t \boldsymbol{X}^-(s)\text{diag}(dN(s))(\boldsymbol{X}^-(s))^T,$$

which is mostly used because of its simple implementation. It is the optional variation process of the martingale $\int_0^t \boldsymbol{X}^-(s)dM(s)$ and it is uniformly consistent.

3.2.4 Inferential procedures

The pointwise confidence interval for $B_j(t)$ is equivalent to

$$\hat{B}_j^*(t) \pm n^{-1/2} c_{\alpha/2} \hat{\Phi}_{jj}^{1/2}(t), \tag{3.7}$$

where $\hat{\Phi}_{jj}(t)$ is the jth diagonal element of $\hat{\Phi}(t)$, and $c_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. It is useful as a synthetic estimator but, as it can vary strongly depending on which time point is chosen, its use for a statistical test about the entire shape of the cumulative regression coefficients would lead to incorrect conclusions.

The two hypotheses $H_0^{(1)} : \beta_j(t) = 0$ (or $B_j(t) = 0$) and $H_0^{(2)} : \beta_j(t) = \gamma$ (or $B_j(t) = \gamma t$), for all t in the range $[0, \tau]$, are of interest, stating, respectively, the assumption of no effect and the assumption of constant effect of the coefficient β_j. Tests are shown directly for the cumulative regression coefficient $B_j(t)$. In order to explain the final test-statistics, we consider the process

$$\Delta_1(t) = n^{-1/2} \sum_1^n \hat{\epsilon}_i^*(t) G_i,$$

which, conditional on the data $(N_i(t), Y_i(t), Z_i(t), X_i(t))$, for $i = 1, \ldots, n$, and under some regularity conditions, has the same limit distribution as $n^{1/2}(\hat{B}^*(t) - B(t))$. The random variables G_i, for $i = 1, \ldots, n$, are independent standard normals and

$$\hat{\epsilon}_i^*(t) = \int_0^t (n^{-1}X^T(s)X(s))^{-1}X_i(s)d\hat{M}_i(s),$$

with

$$\hat{M}_i(t) = \tilde{N}_i(t) - \int_0^t Y_i(s)X_i^T(s)d\hat{B}^*(s).$$

Moreover

$$\hat{\Psi}^*(t) = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^*(t)(\hat{\epsilon}_i^*(t))^T,$$

is a consistent estimator of the asymptotic variance of $n^{1/2}(\hat{B}^*(t) - B(t))$.

The hypothesis $H_0^{(1)}$ can thus be tested using the variance weighted test statistic,

$$T_{1S} = \text{sup}_{t\in[0,\tau]}\left|\frac{n^{1/2}\hat{B}_j^*(t)}{(\hat{\Psi}_{jj}^*(t))^{1/2}}\right|, \tag{3.8}$$

based on the resampling approach for the additive Aalen model by Scheike (2002), where $\hat{\Psi}_{jj}^*(t)$ is the jth diagonal element of $\hat{\Psi}^*(t)$. Since T_{1S} has the same asymptotic distribution as $\sup\left|\Delta_1(t)/(\hat{\Psi}_{jj}^*(t))^{1/2}\right|$ under the null hypothesis, an empirical distribution of this latter can be used to build confidence band for T_{1S}. The empirical distribution is obtained by resampling $\Delta_1(t)$, by generating replicates from the standard normal $\{G_i\}_{i=1,\ldots,n}$. The observed test process can be plotted versus time together with its confidence band. Graphically, $H_0^{(1)}$ may be tested by observing whether the zero function, representing the null hypothesis, is contained within this confidence band.

In order to test the hypothesis $H_0^{(2)}$, the quantity $\hat{B}_j^*(\tau)/\tau$ may estimate the constant γ of the null hypothesis. The two test statistics based on the resampling approach (Scheike, 2002) are :

$$T_{2S} = n^{1/2}\sup_{t\in[0,\tau]}|\hat{B}_j^*(t) - \hat{B}_j^*(\tau)\frac{t}{\tau}| \quad \text{and} \quad T_{2I} = n\int_0^\tau (\hat{B}_j^*(t) - \hat{B}_j^*(\tau)\frac{t}{\tau})^2 dt. \tag{3.9}$$

Approximate p-values can be obtained by resampling from the process $\Delta_1(t) - \Delta_1(\tau)\frac{t}{\tau}$, similar to what was explained earlier. The resampled processes may be plotted versus time, together with the observed process. The possible deviation of this latter from the resampled processes might show rejection of the hypothesis about constant effect.

Note that these test statistics depend on the selected time interval $[0, \tau]$, and therefore different results may be obtained on smaller time intervals.

3.3 The semiparametric additive excess hazards model

The semiparametric additive model is a submodel of the nonparametric additive model where some effects are allowed to be constant in time. We can specify a semiparametric model for relative survival with additive hazards

$$\lambda(t) = Y(t)\{\alpha^*(t; Z) + X^T(t)\beta(t) + V^T(t)\gamma\}, \tag{3.10}$$

where $X(t)$ and $V(t)$ are, respectively, p-dimensional and q-dimensional covariates, $Y(t)$ is the risk indicator, $\beta(t)$ is the p-dimensional time-varying regression coefficient and γ is the q-dimensional time-invariant coefficient. After having tested whether effects are time-varying or constant in the full additive model (3.2), the semiparametric additive model (3.10) could be fitted to better describe the right form of the regression coefficients. Moreover, the model is simpler and leads to less complicated estimators.

3.3.1 Estimators and their properties

The estimators of the cumulative coefficient $B(t) = \int_0^t \beta(s)ds$ and of γ can be obtained by least squares methods, as for the nonparametric additive model. We consider the same setting as for the additive excess hazards model (3.2), where the counting process $N(t)$ is now associated with the intensity $\lambda(t)$ modelled by the semiparametric regression in (3.10). In addition to $\boldsymbol{X}(t)$, define the matrix $\boldsymbol{V}(t) = (Y_1(t)V_1(t), ..., Y_n(t)V_n(t))^T$ of dimension $n \times q$. If we consider the martingale decomposition and the modified counting process $\tilde{N}(t) = N(t) - \Lambda^*(t)$, its corresponding increment can be written as

$$d\tilde{N}(t) = dN(t) - \lambda^*(t)dt = \boldsymbol{X}(t)\beta(t)dt + \boldsymbol{V}(t)\gamma dt + dM(t).$$

Since the martingale increments $dM(t)$ are uncorrelated and with zero mean, least squares methods lead to the equation $d\hat{B}^*(t) = \boldsymbol{X}^-(t)\left(d\tilde{N}(t) - \boldsymbol{V}(t)\gamma dt\right)$ for $B(t)$, where γ has been fixed. The estimator of γ is

$$\hat{\gamma}^* = \left(\int_0^\tau \boldsymbol{V}^T(t)H(t)\boldsymbol{V}(t)dt\right)^{-1} \int_0^\tau \boldsymbol{V}^T(t)H(t)d\tilde{N}(t), \tag{3.11}$$

where the inverse of the matrix $H(t) = W(t)(I - \boldsymbol{X}(t)\boldsymbol{X}^{-}(t))$ is assumed to exist. Finally, plugging the estimator $\hat{\gamma}^*$ into the previous expression for $d\hat{B}^*(t)$, the estimator of $B(t)$ is given as

$$\hat{B}^*(t) = \int_0^t \boldsymbol{X}^{-}(s) \left(d\tilde{N}(s) - \boldsymbol{V}(s)\hat{\gamma}^* ds \right). \tag{3.12}$$

This estimator can also be written as

$$\hat{B}^*(t) = \hat{B}'(t) - \int_0^t \boldsymbol{X}^{-}(s)\hat{\lambda}^*(s)ds + \int_0^t \boldsymbol{X}^{-}(s)\boldsymbol{V}(s) \left[\hat{\gamma}' - \hat{\gamma}^* \right] ds,$$

depending on the estimated p-dimensional cumulative coefficient,

$$\hat{B}'(t) = \textstyle\int_0^t \boldsymbol{X}^{-}(s) \left[dN(s) - \boldsymbol{V}(s)\hat{\gamma}' ds \right],$$

and on the estimated constant coefficient $\hat{\gamma}'$ in a standard semiparametric additive hazards model. The estimator $\hat{\gamma}'$ has the same expression as $\hat{\gamma}^*$ in (3.11), except for the presence of $N(t)$ instead of $\tilde{N}(t)$.

Asymptotic properties of the estimators $\hat{B}^*(t)$ and $\hat{\gamma}^*$ are of primary importance in testing hypotheses about $B(t)$ and γ. Under some regularity conditions, as $n \to \infty$, $n^{1/2}(\hat{\gamma}^* - \gamma)$ converges in distribution to a zero-mean normal V with variance Σ, and $n^{1/2}(\hat{B}^* - B)$ converges in distribution to a zero-mean Gaussian process $U(t)$ with variance $\Phi(t)$. Consistent estimators of the variances Σ and $\Phi(t)$ arise from properties of martingales and the optional variation processes, and they have the same form as for the standard semiparametric additive model (Martinussen and Scheike, 2006).

3.3.2 The maximum likelihood approach

For the semiparametric excess additive hazards model an approximate maximum likelihood estimator can be found, similarly to the one derived by McKeague and Sasieni (1994) for the standard semiparametric model. The partial log-likelihood function can

be written in counting process notation as follows

$$\sum_{i=1}^{n}\left(\int \log\left(\lambda_i(t)\right) dN_i(t) - \int \lambda_i(t)dt\right) =$$
$$\sum_{i=1}^{n}\left(\int \log\left[\lambda_i^*(t) + Y_i(t)X_i^T(t)\beta(t) + Y_i(t)V_i^T(t)\gamma\right] dN_i(t)\right.$$
$$\left. - \int \left[\lambda_i^*(t) + Y_i(t)X_i^T(t)\beta(t) + Y_i(t)V_i^T(t)\gamma\right] dt\right).$$

Derivatives with respect to $\beta(t)$ and γ lead to the score equations

$$\boldsymbol{X}^T(t)\text{diag}(Y_i(t)/\lambda_i(t))\left[d\tilde{N}(t) - \boldsymbol{X}(t)dB(t) - \boldsymbol{V}(t)\gamma dt\right] = 0,$$

$$\int_0^\tau \boldsymbol{V}^T(t)\text{diag}(Y_i(t)/\lambda_i(t))\left[d\tilde{N}(t) - \boldsymbol{X}(t)dB(t) - \boldsymbol{V}(t)\gamma dt\right] = 0,$$

which have the same form as the least squares ones in case that the weight matrix $W(t) = \text{diag}(Y_i(t)/\lambda_i(t))$ and $\lambda_i(t)$ is assumed known. The maximum likelihood estimators with consistent estimates of the weights are asymptotically efficient, as it is in case of the same estimators for the corresponding standard semiparametric model.

3.3.3 Inferential procedures

In order to test the hypothesis of no effect ($\text{H}_0^{(1)}: \ B_j(t) \equiv 0$) and the hypothesis about an effect being time-constant ($\text{H}_0^{(2)}: \ B_j(t) \equiv \gamma_j t$), we suggest to use the confidence band for $B_j(t)$ based on the resampling approach, similarly to what was presented in Section 3.2.4.

From the properties about asymptotic convergence previously described, in the simple case of $W = I$ it follows that

$$\Delta_2(t) = C_1^{-1}n^{-1/2}\sum_1^n \hat{\epsilon}_{2i}^* G_i, \qquad \Delta_3(t) = n^{-1/2}\sum_1^n \hat{\epsilon}_{3i}^*(t) G_i,$$

with $G_1, \ldots, G_n$ independent standard normal, have the same asymptotic distribution as $n^{1/2}(\hat{\gamma}^* - \gamma)$ and $n^{1/2}(\hat{B}^* - B)$ respectively. Then, the variances of these latter are consistently estimated, respectively, by

$$\hat{\Sigma}' = C_1^{-1}(n^{-1}\sum_1^n \hat{\epsilon}_{2i}^*(\hat{\epsilon}_{2i}^*)^T)C_1^{-1}, \qquad \hat{\Psi}'(t) = n^{-1}\sum_1^n \hat{\epsilon}_{3i}^*(\hat{\epsilon}_{3i}^*)^T, \tag{3.13}$$

with

$$\hat{\epsilon}^*_{2i} = \int_0^\tau \left\{ V_i(t) - (V(t)^T X(t))(X(t)^T X(t))^{-1} X_i(t) \right\} d\hat{M}_i(t),$$

$$\hat{\epsilon}^*_{3i}(t) = \hat{\epsilon}^*_{4i}(t) - P(t)C_1^{-1}\hat{\epsilon}^*_{2i}, \quad \hat{\epsilon}^*_{4i}(t) = \int_0^t \left(n^{-1} X^T(s)X(s) \right)^{-1} X_i(s) d\hat{M}_i(s).$$

Vectors $P(t)$ and C_1 are predictable functions of the matrices V and X^- and they are defined as follows:

$$C_1 = n^{-1} \int_0^\tau V^T(t)H(t)V(t)dt, \quad P(t) = \int_0^t X^-(s)V(s)ds.$$

The estimates $\hat{M}_i(t)$ of martingale residuals are

$$\hat{M}_i(t) = \tilde{N}_i(t) - \int_0^t Y_i(s)(X_i^T(s)d\hat{B}^*(s) + V_i^T(s)\hat{\gamma}^* ds).$$

Then, as for the nonparametric excess hazards model, a test statistic for $\mathrm{H}_0^{(1)}$ is

$$T_{1S} = \sup_{t\in[0,\tau]} \left| \frac{n^{1/2}\hat{B}^*_j(t)}{\hat{\Psi}^{1/2}_{jj}(t)} \right|, \tag{3.14}$$

where $\hat{\Psi}_{jj}(t)$ is the jth diagonal element of the robust estimator $\hat{\Psi}(t)$ in equation (3.13). The confidence band for T_{1S} is built resampling Δ_3 in order to find the empirical distribution of $\sup_{t\in[0,\tau]} \left| n^{1/2}\Delta_3(t)/\hat{\Psi}^{1/2}_{jj}(t) \right|$, which has the same asymptotic distribution as T_{1S} under the null hypothesis.

Similarly, for the hypothesis $\mathrm{H}_0^{(2)}$ the test statistics

$$T_{2S} = n^{1/2} \sup_{t\in[0,\tau]} \left| \hat{B}^*_j(t) - \hat{B}^*_j(\tau)\frac{t}{\tau} \right|, \quad T_{2I} = n \int_0^\tau (\hat{B}^*_j(t) - \hat{B}^*_j(\tau)\frac{t}{\tau})^2 dt \tag{3.15}$$

and their quantiles can be computed by resampling from the process $\Delta_3(t)$. Graphical comparisons between the observed test-process $\hat{B}^*_j(t) - \hat{B}^*_j(\tau)\frac{t}{\tau}$ and the simulated processes under the null can show possible time intervals where there is a departure from the hypothesis $\mathrm{H}_0^{(2}$.

3.4 Application to the TRACE data

Data from the TRACE study are here illustrated. They provide a typical example of data exhibiting nonproportional excess hazards with respect to some covariates.

3.4.1 Description of the data

The TRACE study (Kober et al., 1995), consisted in a cohort of 6676 patients with acute myocardial infarction who were screened in 27 Danish coronary care units for entry between May 1990 and July 1992. Information on all patients survival was available from the Danish national registries. The follow-up period was from the day of diagnosis and onwards, during which the outcome under study was total death. The aim of the TRACE study group was to establish which risk factors had a prognostic importance on mortality of patients with acute myocardial infarction.

The actual data set analyzed in this section consists in a random sample of 1876 patients from the TRACE data. Models were fitted only in the follow-up period of the first six years from diagnosis, as most of the excess deaths for myocardial infarction occurred within this time. Patients still alive after six years were considered right-censored. The total number of deaths after myocardial infarction during the follow-up period was 881, and of these, 221 took place within the first two months. The time scale was time since prognosis. The background control population mortality was obtained from the registry StatBank Denmark (www.statistikbanken.dk) during the five years period from 1986 to 1990. Information on the background mortality rates was collected by gender and age.

In our analysis, only the most relevant prognostic factors are taken into account as an example for fitting and comparing the different models. The recorded risk factors are age of patients during the follow-up time, gender (female=1), clinical heart pump failure (CHF) (presence=1), diabetes (presence=1) and ventricular fibrillation (VF) (presence=1). Some risk factors are expected to have effects varying strongly in time, in particular ventricular fibrillation. Previous studies (Jensen et al., 1997) showed that ventricular fibrillation was a very important risk factor for death due to myocardial infarction during the first short time period after diagnosis, but its adverse effect was exhausted approximately two months after.

3.4.2 Comparison of models and estimators

In this section, the nonparametric and semiparametric additive excess hazards models described in Sections 3.2 and 3.3 are analyzed on the TRACE dataset and compared to the standard methods used for modeling relative survival. The total hazard will be written as the sum of the known background rate of mortality in the control population and the excess hazard associated with myocardial infarction.

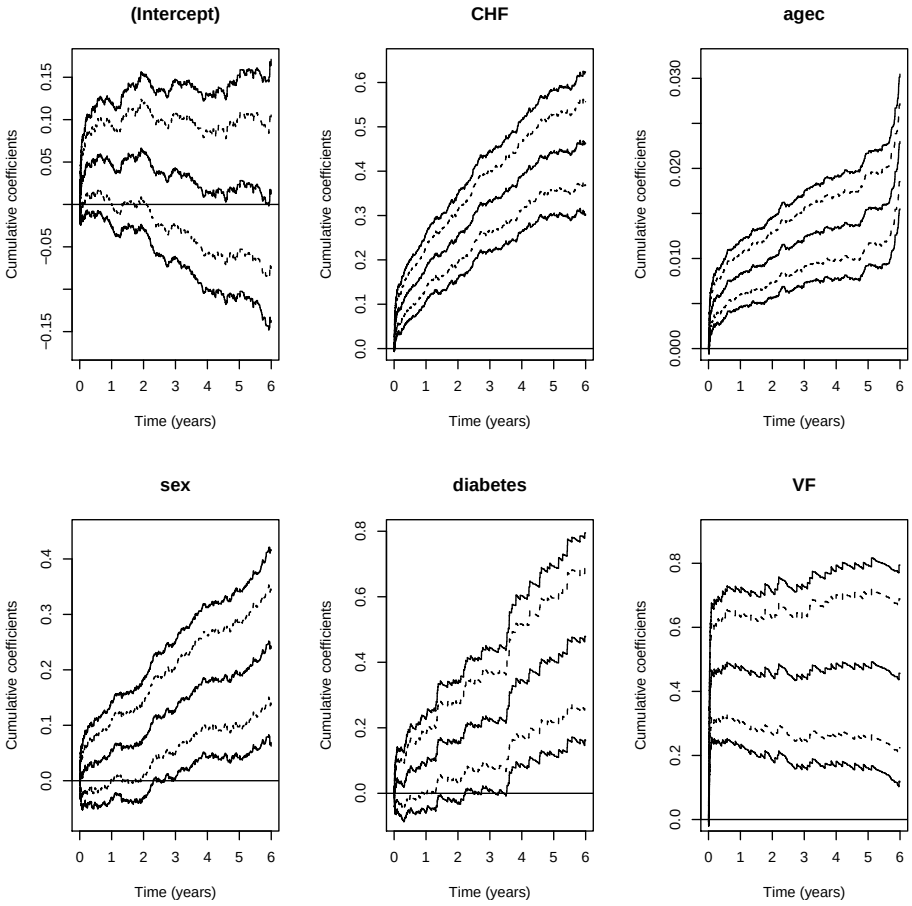

Figure 3.1: *Estimated cumulative regression coefficients for the nonparametric additive hazards model, together with 95% confidence intervals (dashed lines) and confidence bands based on 50 simulated processes under the null (solid lines).*

In the first step, the nonparametric additive excess hazards model is applied to the TRACE data. Successively, we show how possible simplifications of the nonparametric model lead to the more parsimonious semiparametric hazards model. Excess risk for the TRACE data was also estimated through the proportional excess hazards model presented briefly in Equation (3.16) and described more thoroughly in Chapter 4.

Age was centered around its mean at the start of the study (defined as $\bar{a}_0$) and considered as a time-dependent covariate. Results from the nonparametric additive excess hazards model are presented in Table 3.1 and Figure 3.1. For simulation-based tests, a number equal to 300 resampled processes was used. All covariates in the model had an effect significantly different from zero, according to the test T_{1S} in (3.8). Using the supremum test T_{2S} in (3.9), the effects of CHF, centered age and VF resulted to be time-varying, while the effects of gender and diabetes turned out to be invariant

	Test for non-significant effects		Test for time invariant effects			
Covariate	T_{1S}	p-value	T_{2S}	p-value	T_{2I}	p-value
Intercept	2.74	0.147	0.061	0.060	0.006	0.073
CHF	10.20	<0.001	0.121	<0.001	0.043	<0.001
agec	9.76	<0.001	0.006	0.003	1e-04	0.003
sex (female=1)	4.67	<0.001	0.031	0.867	0.001	0.840
diabetes	4.89	<0.001	0.066	0.763	0.003	0.890
VF	6.43	<0.001	0.459	<0.001	0.433	<0.001

Table 3.1: *Nonparametric additive excess hazards model: 300 simulation-based tests for non-significant effects and for time invariant effects.*

in time (Table 3.1). The same conclusions hold in case of using the alternative test statistics T_{2I}. The estimated cumulative regression coefficients $\hat{B}^*(t)$ are shown in Figure 3.1 for each covariate, together with the 95% pointwise confidence intervals (3.7) and the confidence band based on T_{1S} obtained by the resampling technique in Section 3.2.4. The regression function estimates $\hat{\beta}(t)$ are the slopes of the cumulative estimates. Interpretation of their patterns is explained later on in this Section.

Particular care needs to be taken in the interpretation of the intercept $\beta_0(t)$ and its behaviour in the model when compared with the horizontal zero line. In our application, the excess intensity for a male subject without CHF, diabetes and VF, is $\nu_i(t) = Y_i(t)\,[\beta_0(t) + ((a_{0i} + t) - \bar{a}_0)\beta_1(t)]$, where a_{0i} is the age of subject i at the start of the study. In this case, the intercept needs to be interpreted together with the additional coefficient β_1. The excess baseline hazard can then be represented by $\nu_i(t)$ for a subject with $a_{0i} = \bar{a}_0$. In order to interpret correctly the coefficient $\beta_0(t)$ on its own as the excess baseline hazard, the time-dependent age $a_{0i} + t$ should be centered with respect to $\bar{a}_0 + t$. Thus, the additional term about age in the excess hazard would be null for a subject with mean age $\bar{a}_0 + t$ at every t. In this second case, results from the application (not shown here) indicated that patients with acute myocardial infarction have an estimated decreasing relative survival during the whole follow-up period.

From Figure 3.1 it can be observed that the effects of gender and diabetes on the excess mortality rate are constant with time, as graphs of their estimated cumulative coefficients are approximately straight lines. The time invariance of these two covariate effects justifies a possible simplification of the model by reducing the number of nonparametric components. Therefore, the semiparametric additive excess hazards model is also applied to the TRACE data, where effects of gender and diabetes are assumed to be constant and the remaining covariate effects are allowed to be time-varying.

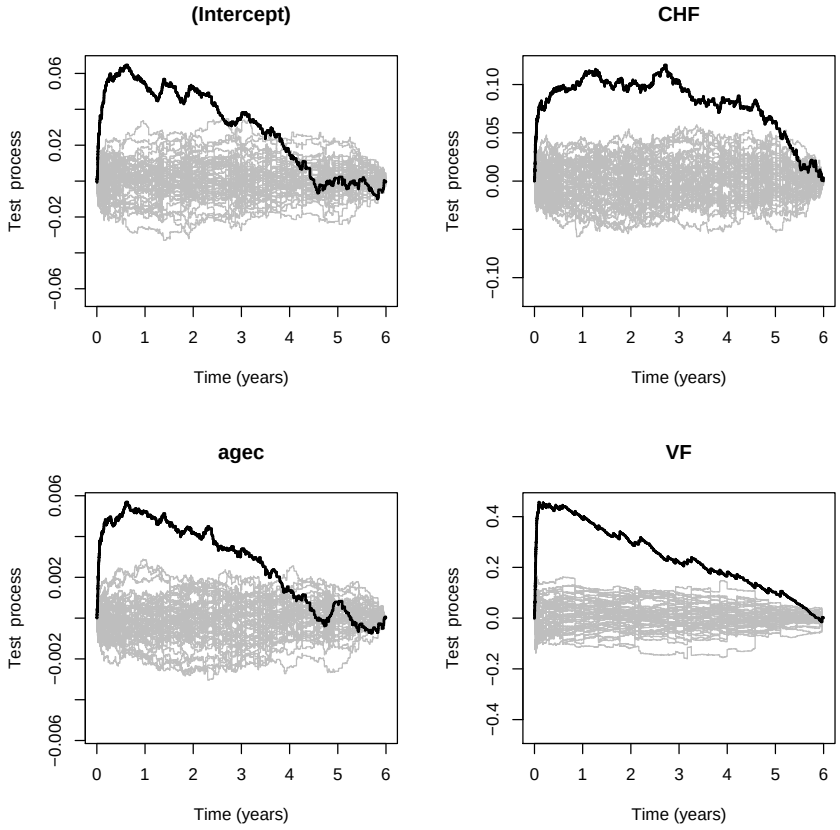

Figure 3.2: *Semiparametric additive excess hazards model: Observed test-process for each covariate, along with 50 simulated processes under the null hypothesis of time invariant effects.*

Results about the semiparametric excess hazards model are presented in Table 3.2. The assumption of constant effects for gender and diabetes was confirmed by the results in the right side of Table 3.2. According to the tests T_{2S} and T_{2I} in (3.15), the remaining covariate effects were still significantly time-varying, as in the previous nonparametric model. This reduced semiparametric model gives a better fit to the TRACE data, as it is simpler in the interpretation and able to discriminate between constant and time-varying effects. Moreover, going from the nonparametric to the semiparametric additive model, comparison of Table 3.1 with Table 3.2 reveals that values of the supremum and squared tests are almost unchanged. Graphics about behaviour of the estimated time-varying cumulative coefficients were also unchanged with respect to the nonparametric case, and thus they can be observed from the same Figure 3.1. Both the constant effects in Table 3.2 were significant (p-values < 0.001) and positive: For patients with diabetes the estimated excess mortality rate was 8.3% higher than for patients without diabetes and this increase was estimated to be constant within the 6

Test for time invariant effects					Constant effects		
Covariate	T_{2S}	p-value	T_{2I}	p-value	Covariate	Coef. γ	Robust SE
Intercept	0.065	0.003	0.008	0.003	sex (female=1)	0.043	0.010
CHF	0.120	<0.001	0.043	<0.001	diabetes	0.083	0.018
agec	0.006	0.003	0.0001	0.003			
VF	0.458	<0.001	0.432	<0.001			

Table 3.2: *Semiparametric additive excess hazards model: 300 simulation-based tests for for time invariant effects and estimates of constant effects.*

years follow-up; the female gender was associated with an estimated increased excess mortality rate of 4.3%. Departure of the effects from the null hypothesis of time invariance may be observed easily looking at Figure 3.2, where each observed test-process is shown along with 50 resampled processes under the null. Presence of a significant variation within the six years follow-up period is very evident for the VF coefficient: Behaviour of its test-process in Figure 3.2 reveals that the effect of ventricular fibrillation is very strong initially, and thus the excess mortality rate has a very high increase within the first two months, but successively the effect seems to disappear in time. Increasing age had also a strongly time-varying effect, which was very high within approximately the first eight months. Similarly, the effect of CHF was increasing very fast initially, after two months it continued to be present but constant until the fourth year, finally the effect vanished during the last two years of follow-up.

We apply the proportional excess hazards model by Sasieni (1996) to the TRACE data, in order to verify whether the excess hazard associated with myocardial infarction could be described by a proportional form. The statistical model is

$$\lambda(t) = Y(t) \left[\alpha^*(t; z) + \lambda_0(t)\exp(X^T\beta)\right], \tag{3.16}$$

where the regression coefficient β is assumed to be time invariant. A formal description of this model can be found in Section 4.1.1. The same set of covariates analyzed in the previous models is influencing significantly the proportional excess hazard, by increasing it (Table 3.3). CHF and VF seem to be very important risk factors in predicting the excess mortality rate due to myocardial infarction, as for patients with heart pump failure or with ventricular fibrillation the excess hazard ratio is about 3.2 and 2.7, respectively. Nevertheless, these last results could be questionable because they are related to regression coefficients which are assumed to be invariant in time. If instead effects of CHF and VF were highly time-varying (as it was in the semiparametric additive model), the assumption of proportional excess hazards would be violated, since

Test for non-significant effects				
Covariate	$\exp(\beta)$ (Rel. risk)	SE(β)	95% CI for rel. risk	p-value
CHF	3.158	0.130	(2.436 - 4.056)	<0.001
agec	1.046	0.005	(1.035 - 1.057)	<0.001
sex (female=1)	1.689	0.117	(1.342 - 2.125)	<0.001
diabetes	1.998	0.120	(1.579 - 2.529)	<0.001
VF	2.718	0.131	(2.109 - 3.522)	<0.001

Table 3.3: *Proportional excess hazards model: Tests for non-significant effects.*

it is strictly related to the invariance of the regression coefficients in the relative risk.

3.5 Discussion

The high flexibility of the additive nonparametric and semiparametric models for relative survival, together with the inferential aspects described in this Chapter, provides a very important alternative to the existing methods in this field, and on the other hand, a useful general extension of the more restrictive recent models. Indeed, the model fitting may fail both because the chosen link function for the excess hazard (multiplicative or additive function) is inappropriate, and because the time invariance of the hazard ratio does not hold, besides misspecification of the functional forms of covariates. The described additive excess hazards models overcome the critical problem of violating the proportional hazards assumption. The introduction of covariate-by-time interactions in the parametric relative survival models entails further assumptions which would need always to be carefully tested, in order to avoid neglecting possible associations between time-dependent covariates and excess mortality.

The TRACE example demonstrates the need of new flexible survival models for modeling the excess hazards, which can deal with time-varying dynamics of covariates effects. In this Chapter, we showed how the nonparametric and semiparametric versions of the additive excess hazard can easily handle these dynamics. We demonstrated when one or the other model is appropriate according to the responses of simulation based graphical and statistical tests about variation of effects over time. Even though inferential procedures described here are complicated in their expressions, when they concern finding equivalent asymptotic distributions of Gaussian processes, the great advantage is a very easy interpretation of results. In this connection, the statistical software, e.g. the R package `timereg` (Martinussen and Scheike, 2006, App. C) used in our application and presented in the Appendix A, is an essential instrument.

The graphical procedures showed for the additive excess hazards models have the advantage of suggesting time points and sub-intervals where variation of the effects occurs in time with sufficient accuracy, while, in the graphics about Brownian bridge processes (Stare et al., 2005) for the proportional excess models, these information are not clearly provided because of the implementation of smoothing procedures.

It would be of interest to extend other test methods about time-varying covariate effects and goodness-of-fit plots from the nonparametric and semiparametric additive hazards model to the relative survival case. Some starting points could be Aalen (1989, 1993) and Gandy et al. (2007).

As for the nonparametric excess additive hazards model, also for the semiparametric model (3.10), approximate maximum likelihood estimators can be found, similarly to what was done by McKeague and Sasieni (1994) for the semiparametric additive hazards models. They are also asymptotically efficient in case of consistent estimates of the weights. For the model by McKeague and Sasieni (1994), there exist also other estimators, improved by their properties of robustness and consistency, which could be easily extended to the relative survival case when the replacement of $d\tilde{N}(s)$ holds.

In choosing between a proportional or an additive form for the excess hazards, problems about non-proportionality and large number of covariates under study should always be faced. An additional crucial problem, which was not studied in the present Chapter, concerns non-positive excess hazards in relative survival regression models. From a practical point of view, models about some situations as prevention studies, would need to allow the excess hazards to be negative, assuring however non-negative observed intensities. As pointed by Zahl (1996), the nonparametric additive excess hazards model overcomes this problem. A proportional excess hazards model can not be used in case of negative excess intensities, however, it is still possible to consider a possible excess intensity equal to zero (Sasieni, 1996).

Models which allow accommodating time-varying covariate effects are very interesting in the relative survival scenario. In this chapter, we studied such models with additive excess hazards and presented the usefulness of some inferential procedures about time-varying coefficients. A natural and important case to investigate, following the same lines of study presented here, would be allowing the presence of both time-varying and constant regression coefficients within the proportional excess hazards model (3.16).

Chapter 4

Goodness-of-fit for Relative Survival Models

The purpose of this chapter is to describe a suggestion for goodness-of-fit methods and graphical tests for residuals in the relative survival setting. We do this by a straightforward use of the cumulative martingale residuals proposed by Lin et al. (1993), and we illustrate how to use the cumulative martingale residuals for testing the proportional hazards assumption in the proportional excess model by Sasieni (1996). This approach is very simple to implement and is known to work well in the standard survival setting.

4.1 Introduction and background

There is a general lack of accomplished methodology for regression diagnostics and assessment of goodness-of-fit of additive relative survival models. The existing theory is only sometimes implemented in public software. Some of the parametric models are estimated in the framework of generalized linear models, thereby enabling the use of standard regression diagnostics in this area. Recently, in the context of models with multiplicative excess rate, Stare et al. (2005) proposed some diagnostics aimed at detecting time-varying effects of covariates on the excess risk and based on partial residuals defined similarly to the Schoenfeld residuals for Cox model. However, their procedure relies heavily on the choice of a smoothing parameter, that can be completely avoided by the procedure we suggest here. An additional problem with the proposal by Stare et al. (2005) is that it does not lead to the correct level even though it in practice tend to work well.

In order to show our idea, we need some background which was already partially introduced in Chapter 3. We describe the proportional excess hazards model, one of the semiparametric choices in relative survival. Martingale residuals and their properties are already reviewed in Section 2.1.2. The successive sections concern the application of cumulative martingale residuals, presented in Section 2.1.2, to the proportional excess hazards model and to the nonparametric additive excess hazards model described in Section 3.2.

4.1.1 The proportional excess hazards model

The proportional excess model proposed by Sasieni (1996) models the excess risk on a multiplicative scale. The statistical model is

$$\lambda(t) = Y(t)\left[\alpha^*(t;z) + \lambda_0(t)\text{exp}(X^T\beta)\right], \tag{4.1}$$

where the p-dimensional regression coefficient $\beta = (\beta_1, \ldots, \beta_p)^T$ is assumed time invariant. The p-dimensional vector X contains the covariate values. Here the notation is as in Section 3.2. In the counting process setting, the intensity $\lambda(t)$ is associated with the process $N(t)$, with $t \in [0, \tau]$, $\tau < \infty$. Referring to the same definitions of the model in (3.2), we associate the compensated counting process $\tilde{N}(t) = N(t) - \Lambda^*(t)$ to the martingale $M(t)$ so that

$$M(t) = \tilde{N}(t) - \int_0^t Y(s)\text{exp}(X^T\beta)d\Lambda_0(s),$$

with $Y(t) = (Y_1(t), \ldots, Y_n(t))$.

Solving the unweighted score equations derived from the log-likelihood for β and λ_0, up to all the observation period $[0, \tau]$, leads to the following estimator for the baseline cumulative excess hazard

$$\hat{\Lambda}_0(t;\beta) = \int_0^t \frac{\sum d\tilde{N}_i(u)}{\sum Y_i(u)\text{e}^{X_i^T\beta}}, \tag{4.2}$$

with X_i equal to the p-dimensional vector of covariates of subject i. The substitution of this estimator in the score equation for β yields

$$U(\beta) = \sum_i \int \left\{ X_i - \frac{\sum_j Y_j(t)X_j\text{e}^{X_j^T\beta}}{\sum_j Y_j(t)\text{e}^{X_j^T\beta}} \right\} d\tilde{N}_i(t), \tag{4.3}$$

which provides an estimate for the parameter β such that $U(\hat{\beta}) = 0$. The general background theory for the score equations is given in 1.3.4. Setting $\lambda_i^*(t) = 0$ for $i = 1, \ldots, n$, the modified counting process $\tilde{N}_i$ is equal to N_i and the unweighted estimators are the solutions to the usual score equations for the Cox model (see equation (1.32) and Example 1.3.2 for more details). Properties of the estimators and conditions under which they are valid can be found in Sasieni (1996).

Note that some difficulties arise in exchanging summation with integration in equation (4.3), which depends on both the observed failure times and the observed censoring times, as the modified counting process $\tilde{N}_i$ changes at every censoring time, besides at every failure time.

4.2 Goodness-of-fit with cumulative martingale residuals

In this section, we propose a very straightforward procedure based on cumulative martingale residuals for testing goodness-of-fit of the proportional excess hazards model (4.1). Our approach can also be used to assess the fit of the additive hazards excess model but we here illustrate the basic idea by looking at the proportional excess model.

In the proportional excess hazards model, we are interested in checking whether the sub-model for the excess hazard is adequate. More specifically, in order to fulfill this objective, three aspects would need to be checked: Functional form of covariates, the form of the link function of the excess hazard, the assumption of proportional hazards. We show how the cumulative sums of martingale-based residuals (Lin et al., 1993) can be used to answer this problems.

The partial likelihood score function (4.3) for the parameter β can be also written as a functional of the martingale process $M_i(t)$ associated with individual i, as in equation (2.11). Here we recall the expression of the score function up to the entire interval $[0, \tau]$,

$$U(\beta) = \sum_i \int_0^\tau \{X_i - E(\beta, t)\}\, dM_i(t),$$

where we have defined $E(\beta, t) = S_1(\beta, t)/S_0(\beta, t)$ and

$$S_k(\beta, t) = \sum_i Y_i(t) X_i^{\otimes k} \exp(X_i^T \beta),$$

with $k = 0, 1, 2$. We have $X_i^{\otimes 0} = 1$, $X_i^{\otimes 1} = X_i$ and $X_i^{\otimes 2} = X_i X_i^T$. $X_i = (X_{1i}, \ldots, X_{pi})^T$ is the p-dimensional vector of covariates of individual i.

The martingale residuals (Section 2.1.2) for the proportional excess hazards model are defined as

$$\hat{M}_i(t) = \tilde{N}_i(t) - \int_0^t Y_i(s)\exp(X_i^T\hat{\beta})d\hat{\Lambda}_0(s), \tag{4.4}$$

where $\hat{\Lambda}_0(s)$ is the estimator in (4.2). They are defined similarly to the martingale residuals for the standard proportional hazards model (Grønnesby and Borgan, 1996). They verify the basic properties given in Section 2.1.2, i.e., their sum over the individuals is zero and they average to zero asymptotically. The cumulative martingale residuals (Section 2.1.2) are constructed by different partial-sum processes of the martingale residuals $\hat{M}_i(t)$. Processes can be over follow-up time or covariate values, in order to test, respectively, the proportional excess hazards assumption or the functional form of covariates and the link function. Then, tests about these aspects are made by using the processes to compare their observed behaviour with their potential one under the assumption that the model is true.

The functional of the martingale residuals used to test the proportional excess hazards assumption is based on the observed score process in time, written as $U(\hat{\beta}, t) = \sum_i X_i\hat{M}_i(t)$. Using the cumulative martingale residuals $U_j(\hat{\beta}, t) = \sum_i X_{ji}\hat{M}_i(t)$, the proportional excess hazard assumption may be verified both by graphical plots and by hypothesis tests. A test statistics for each j ($j = 1, \ldots, p$) is given by the supremum of the standardized score process

$$\sup_t \left(\widehat{Var}(U_j(\hat{\beta}, t)\right)^{-\frac{1}{2}} \left|U_j(\hat{\beta}, t)\right|, \tag{4.5}$$

where

$$\widehat{Var}\left(U_j(\hat{\beta}, t)\right) = \sum_i \int_0^t \left(X_i - \frac{S_1(\hat{\beta}, s)}{S_0(\hat{\beta}, s)}\right)^{\otimes 2} dN_i(s)$$

is a consistent estimator of the variance of the observed score process. This supremum test for proportionality has the advantage that no specific functional form needs to be chosen when looking for lack of fit of the model for a specific covariate j.

The distribution of $n^{-\frac{1}{2}}U(\hat{\beta}, t)$ is asymptotically equivalent to

$$n^{-\frac{1}{2}}\left(\hat{D}_1(t) - \hat{J}(\hat{\beta}, t)\hat{J}^{-1}(\hat{\beta}, \tau)\hat{D}_1(\tau)\right), \tag{4.6}$$

where $\hat{D}_1(t) = \sum_{i=1}^n \int_0^t (X_i - E(\hat{\beta}, s))dN_i(s)G_i$ and $G_1, \ldots, G_n$ are independent standard normals. The matrix $\hat{J}(\hat{\beta}, t)$ represents minus the derivative of the score function with respect to β given by (4.3). Then, the asymptotic distribution of $n^{-\frac{1}{2}}U(\hat{\beta}, t)$

may be evaluated using a resampling procedure, by generating realizations from the process (4.6) which depends only on the random variables G_i (Martinussen and Scheike, 2006). This is made by repeatedly generating normal random samples $\{G_i\}$ while holding the observed data $\{N_i, Y_i, X_i\}$ fixed. The null distribution of the test statistics in (4.5) is then approximated by these simulations. A graphical test about proportionality may be obtained by plotting the observed score process $U(\hat{\beta}, t)$ over time together with the realizations we have simulated from the process (4.6) in order to approximate the null distribution of $U(\hat{\beta}, t)$. If the observed score process diverges from the simulated processes under the model, which should randomly fluctuate around the zero axis, there is evidence of a lacking fit of the proportional excess hazards model due to the missing proportionality.

The key reasoning about the validity of the cumulative martingale residuals in checking the current model consists in replacing the counting process $N(t)$ with the modified counting process $\tilde{N}(t)$ when it is opportune. Here, we underline the use of the estimator $\hat{\Lambda}_0$ in (4.2), expressed as a function of $\tilde{N}(t)$. Moreover, in the process (4.6), the estimator of minus the average of the derivative of the score function with respect to β, $\hat{J}(\hat{\beta}, t)$, needs also to be a function of $\tilde{N}(t)$ and it is evaluated as

$$\hat{J}(\hat{\beta}, t) = \sum_i \int_0^t \left(\frac{S_2(\hat{\beta}, s)}{S_0(\hat{\beta}, s)} - E(\hat{\beta}, s)^{\otimes 2} \right) d\tilde{N}_i(s).$$

Finally, it is important to note that the process (4.6) depends directly only on the original counting process $N_i(t)$, but not on $\tilde{N_i}(t)$, as the variance of $M_i(t)$ is equal to $E(N_i)$, and therefore can be approximated by $G_i N_i$.

Graphical and statistical tests for checking the functional form of covariates and the link function in the proportional excess hazards model may be carried out very similarly to the ones proposed by Lin et al. (1993) for the proportional hazards model and involve the same substitutions shown previously in this section. For investigation of the functional form of a certain covariate j, the tests are based on the cumulative residual process $M_c^j(x) = \sum_{i=1}^n \int_0^\tau I(X_{ji}(t) \leq x) d\hat{M}_i(t)$, where $I(\cdot)$ is the indicator function, $x \in \mathbf{R}$, and $\hat{M}_i(t)$ are defined in (4.4). Resampling methods, as described previously, provide simulated realizations under the null, which approximate the asymptotic distribution of the latter process. Therefore, a graphical test is given by plotting the observed cumulative residuals $M_c^j(x)$ versus the continuous covariate with values x, together with random realizations under the model.

Test for proportionality of the excess hazard		
Covariate j	Test-statistics $\sup_t \lvert U_j(t) \rvert$	p-value
CHF	12.7	0.418
agec	163.0	0.826
sex (female=1)	22.1	0.176
diabetes	19.4	0.086
VF	29.2	0.002

Table 4.1: *Proportional excess hazards model: 50 simulation-based tests for proportionality of the relative excess risk.*

4.3 Example from the TRACE data

In order to test proportionality of the excess hazards of each covariate in the proportional model (4.1), we use the simple non-standardized version of the test-statistics (4.5) based on cumulative martingale residuals. Results in Table 4.1 suggest that only the covariate VF contributes to violate the assumption of proportionality ($p = 0.002$), whereas the proportional effect of CHF was correctly verified by the data.

Comparison of the model-based relative survival functions with the corresponding nonparametric estimated curves, underlines the possible violation of assumptions in the analyzed models. We considered the semiparametric additive excess hazards models and the proportional excess hazards model with sex and VF as the only risk factors. In Figure 4.1, the four estimated relative survival functions from each model are compared with the corresponding relative survival curves (relative survival ratios) estimated by using the Kaplan-Meier method for the observed and the Hakulinen method (Hakulinen, 1982) for the expected survival. The choice of the alternative Edered II method (Ederer and Heise, 1959) for the expected survival does not affect the final results, as our example concerns a short follow-up period. In panel (a) of Figure 4.1, it is observed that the proportional excess hazards model does not fit very well data of patients with ventricular fibrillation, neither for females nor for males. On the other hand, this model captures well the difference in relative survival between males and females. The current lack of fit of the proportional excess hazards model is due to the wrong assumption of proportional excess hazards for VF, which does not reflect a much higher excess risk of dying soon after admission in the study for patients with ventricular fibrillation. Predictions in panel (b) of Figure 4.1 describe much better the excess mortality pattern for the different patients groups, since the presence of a time-varying coefficient for VF in the semiparametric model allows to capture changes of the effect of VF with time.

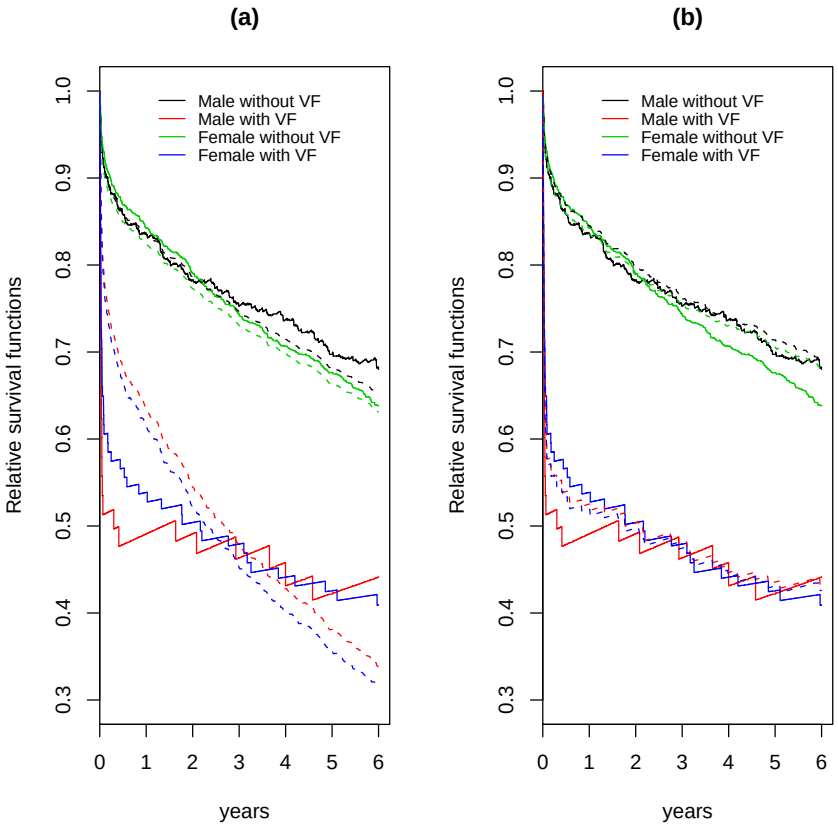

Figure 4.1: *(a) Comparison between relative survival predictions based on the proportional excess hazards model (dashed lines) and nonparametric relative survival estimates based on the Kaplan-Meier and Hakulinen methods (solid lines) by sex and VF. (b) Comparison between relative survival predictions based on the semiparametric excess hazards model (dashed lines) and nonparametric relative survival estimates based on the Kaplan-Meier and Hakulinen methods (solid lines) by sex and VF.*

The evidence of the wrong assumption about the proportionality for the VF effect within the additive model with excess risk as in (4.1), was also provided by the statistical and graphical tests proposed by Stare et al. (2005), based on the maximum values of the Brownian bridge processes. The EM method for smoothed baseline excess hazards was chosen within the R package `relsurv` (Pohar and Stare, 2006), in order to fit the regression model. The effect of VF resulted to be time-varying (maximum value was equal to 3.109 with $p < 0.001$), whereas CHF and all the remaining covariates had time-constant effects. Therefore, the analyses of goodness-of-fit based on the test-statistics (4.5) and on the tests by Stare reached the same conclusions about the TRACE study, stating that it is solely the covariate VF that ruins the proportional effects. Different results given by the models presented in the current section are essentially due to modeling the excess risk on different scales, that is, the proportional or

the additive scale. The effect for CHF, which resulted to be time-varying when using the latter scale in the additive excess hazards models, but time-constant in models with the proportional scale, is an example of that.

4.4 Discussion

Even though our suggestion is related to the recent interesting proposal by Stare et al. (2005), our approach has important advantages. First, our method does not need any critical choice of smoothing parameters (or parametric assumptions) for the baseline. Secondly, our procedure is asymptotically justified and will thus lead to asymptotically correct p-values and this is not true in general for the Stare et al. procedure.

Our suggestions about checking goodness-of-fit of the proportional excess hazards model and the additive excess models play an important role in a good model selection. An advantage of the supremum test described in Section 4.2 is that no specific deviations from proportionality need to be explicitly expressed. The drawback is however that the model is assumed to be correct with respect to all the other covariates when the proportionality assumption is investigated for a specific covariate. Nevertheless, this is a general problem faced also by the existing methods for goodness-of-fit of regression survival models. Then, important features of the data may be overlooked, and we might be unable to detect where a possible lack of proportionality occurs during the follow-up time.

Chapter 5

Outlook: Time-dependent Covariates in Competing Risks Settings

The application to breast cancer in Chapter 2 gave rise to investigating the role of time-dependent covariates in competing risks regression models, and more generally, in multi-state regression models. There exist various types of time-dependent covariates, which differ in their random or deterministic development in time (Appendix B). When some of these are studied, predictions based on the model are not allowed, or they meet with difficulties.

The area of research about the role of time-dependent covariates in this field is at a young stage and there exists little literature focusing on how to handle different types of time-dependent covariates. The present chapter is an attempt to enter this area and provide some directions for future work.

In the next section a general overview of the state of the art, problems and future directions are introduced. The following section presents a possible extension of the competing risks model, that allows us to include a simple random binary time-dependent variable, in a multi-state framework. Inclusion of the sojourn time of an individual in a certain state as a time-dependent covariate into the model, is also studied.

5.1 Introduction

In multi-state models, and specifically in competing risks models, the principal interest focuses often on the cumulative incidence probabilities. When a regression analysis is suitable, in general the aim is to investigate the effects of covariates on these probabilities. Of course, both time-independent and time-dependent covariates may be relevant to study.

The standard approach (Andersen et al., 1993, Chap. 7) consists of separate regression models for all the cause-specific hazards, which are then combined to estimate the cumulative incidence probabilities. The effect of the covariates on these probabilities is not direct and can not be synthesized by simple regression parameters. This difficulty has led to the development of alternative recent approaches, which aim to establish direct effects of covariates on the cumulative incidence probabilities (Scheike and Zhang, 2007). Fine and Gray (1999) proposed the proportional subdistribution hazards model for competing risks, and the direct parametric inference for the cumulative incidence functions is discussed by Jeong and Fine (2006). Klein and Andersen (2005) presented a further approach based on pseudovalues. Predictions of cumulative incidence functions by the direct binomial regression approach are given by Scheike et al. (2007).

Regression on some kinds of time-dependent covariates, especially internal covariates (Appendix B), leads to problems in interpreting and predicting cumulative incidence probabilities within the standard approach, as discussed in the next section. These aspects belong to the class of problems arising when model specification is only partial (Andersen et al., 1993, Chap. 3). A possible solution, which however yields rather complex theory, consists of specifying completely the model, i.e., giving a joint model for the multi-state process and the time-dependent covariates (Henderson et al., 2000). An open question is the role of external and internal covariates (Appendix B) in modelling cumulative incidence probabilities according to the previously mentioned alternative approaches, and whether or not predictions are possible.

The present chapter, within the standard approach of Andersen et al. (1993, Chap. 7), provides a discussion about the role of the different types of time-dependent covariates, throws light on some interesting directions of work and attempts to give some possible solutions. This work could serve as a starting point for further investigation on the above mentioned problems and related aspects, both within the standard approach and within the alternative recent approaches.

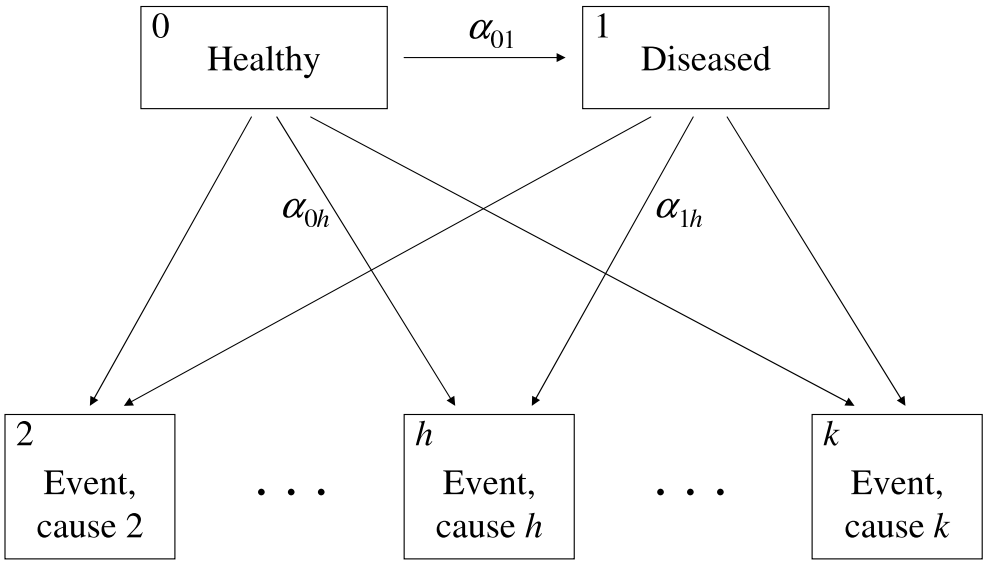

Figure 5.1: *The extended illness-death model for competing risks without possibility of recovery. The ending event can be due to causes* $\{2, \ldots, k\}$*. The arrows represents the possible transitions between states.*

5.2 An extended illness-death model for competing risks

This section presents a multi-state model where additional information about the healthy and diseased states is joint to the standard competing risks model. Therefore, we can think of this model as an extended illness-death model (Appendix C) where the competing risks are also taken into account.

Consider a nonhomogeneous Markov process in continuous time with finite state space $\mathcal{S}_c = \{0, 1, 2, \ldots, k\}$. The extended illness-death model for competing risks has absorbing states $\{2, \ldots, k\}$, representing the ending events due to different causes, while the states 0 and 1 are transient and represent two different statuses of individuals. Its possible transitions are shown in Figure 5.1, where, in clinical studies, the patient status can be 'healthy' or 'diseased'. Individuals can experience the transition from 0 to 1, but the transition back to state 0, i.e., the possibility of recovery for diseased patients is here excluded. Moreover, both 'healthy' and 'diseased' individuals can meet with one of the possible $k - 1$ ending events.

The probability space is $(\Omega, \mathcal{Z}, P)$, with $\mathcal{Z}$ being the filtration generated by the stochastic process. Denote with $P_{hl}(s, t)$ and $\alpha_{hl}(t)$, $h, l \in \mathcal{S}_c$, the transition probabilities and the transition intensities, respectively, as in equations (1.33) and (1.34). Under independent right-censoring, let $\boldsymbol{N}(t) = (N_{0l}(t), l \in \{1, \ldots, k\}; N_{1h}(t), h \in \{2, \ldots, k\})$ be the multivariate counting process, where, for instance, $N_{0l}(t)$ counts the number of

direct observed transitions from 0 to l in $[0, t]$. Assume that $\boldsymbol{N}(\cdot)$ has intensity process $\boldsymbol{\lambda}(t) = (\lambda_{0l}, l \in \{1, \ldots, k\}; \lambda_{1h}(t), h \in \{2, \ldots, k\})$, where each element has the multiplicative form $\lambda_{0l}(t) = Y_0(t)\alpha_{0l}(t)$, with $Y_0(\cdot)$ being the at-risk process. The hazard functions $\alpha_{0l}(\cdot)$, for $l \in \{1, \ldots, k\}$, and $\alpha_{1h}(\cdot)$, for $h \in \{2, \ldots, k\}$, regulate the behaviour of the extended illness-death model for competing risks, as shown by the arrows in Figure 5.1.

Let $\boldsymbol{A}$ be the matrix of integrated transition intensities, where the positive elements are $A_{0l}(t) = \int_0^t \alpha_{0l}(u)du$, for $l \in \{1, \ldots, k\}$, and $A_{1hl}(t) = \int_0^t \alpha_{1h}(u)du$, for $h \in \{2, \ldots, k\}$. $A_{hh}(t) = \int_0^t \alpha_{hh}(u)du$, for $h = 0, 1$, are defined so that $\alpha_{hh}(\cdot) = -\sum_{l \neq h} \alpha_{hl}(\cdot)$. The transition matrix $\boldsymbol{P}$ can be written in product integral representation as in Equation (1.35), with $Q = d\boldsymbol{A}$. Since we are in the absolutely continuous case, $\boldsymbol{P}$ is the unique solution to the Kolmogorov forward differential equations for the intensity matrix $\boldsymbol{\alpha}$. As a solution to these equations, explicit expressions for the transition probabilities of the extended illness-death model for competing risks can easily be obtained.

The probability $P_{01}(s, t)$ is equal to the one in the illness-death model (Appendix C), specified in equation (C.1). Its interpretation is intuitive: an individual may sojourn in state 0 for a certain time $[s, u-]$ with probability $P_{00}(s, u-)$, then he may jump to state 1 at u with instantaneous rate $\alpha_{01}(u)$ and finally sojourn in state 1 the remaining time $[u, t]$ with probability $P_{11}(u, t)$. The transitions from state 1 ('diseased') to the cause-specific events are regulated by the probabilities

$$P_{1h}(s, t) = \int_s^t P_{11}(s, u-)\alpha_{1h}(u)du, \qquad h = 2, \ldots, k, \quad s \leq t. \tag{5.1}$$

The cumulative incidence probabilities for the cause-specific events are the transition probabilities from state 0 to the absorbing states $2, \ldots, k$, given by

$$P_{0h}(s, t) = \int_s^t \left[P_{00}(s, u-)\alpha_{0h}(u) + P_{01}(s, u-)\alpha_{1h}(u)\right] du, \quad h = 2, \ldots, k, \quad s \leq t. \tag{5.2}$$

Equations (C.1), (5.1) and (5.2) depend on the probabilities of permanence in 0 or in 1 between a certain time interval $[s, t]$, $P_{00}(s, t)$ and $P_{11}(s, t)$. Provided that $A_{hh}(\cdot)$, $h = 0, 1$, are absolutely continuous, their product integral representation has the explicit form

$$P_{hh}(s, t) = \exp\left\{\int_s^t \alpha_{hh}(u)du\right\}, \qquad h = 0, 1, \quad s \leq t, \tag{5.3}$$

where

$$\alpha_{00}(u) = -(\alpha_{01}(u) + \sum_{h=2,\dots,k} \alpha_{0h}(u)), \qquad \alpha_{11}(u) = -\sum_{h=2,\dots,k} \alpha_{1h}(u).$$

One of the primary interests is to estimate the $P_{0h}(s,t)$, $h = 2,\dots,k$, which represent the probability of experiencing the event due to cause h within $(s,t]$, taking the history of the disease into account. Another important aspect to study is the marginal survival probability, that is the probability of not experiencing any event until time t, given by

$$S(t) = P_{00}(t) + P_{01}(0,t). \tag{5.4}$$

First, we recall the Nelson-Aalen estimators for the cumulative hazard functions,

$$\hat{A}_{0l}(t) = \int_0^t \frac{dN_{0l}(u)}{Y_0(u)}, \quad \hat{A}_{1h}(t) = \int_0^t \frac{dN_{1h}(u)}{Y_1(u)}, \quad l = 1,2,\dots,k, \quad h = 2,\dots,k, \tag{5.5}$$

which are assumed to be equal to zero at times where the at-risk process is not positive. Moreover, define $\hat{A}_{hh}(t) = -\sum_{l\neq h} \hat{A}_{hl}(t)$, for $h = 0,1$, and denote with $\hat{\boldsymbol{A}}$ the matrix containing the estimators of the cumulative hazard functions. For finite-state Markov processes, an important result consists in the so-called Aalen-Johansen estimator of the transition matrix $\boldsymbol{P}$ (Aalen and Johansen, 1978). This estimator is

$$\hat{P}(s,t) = \prod_{(s,t]} \left(\boldsymbol{I} + d\hat{\boldsymbol{A}}(u) \right). \tag{5.6}$$

It is expressed by the product integral representation similarly to the Kaplan-Meier estimator of the survival probability in a simple two states model, with one of them absorbing.

The fundamental result (5.6) yields the following estimators

$$\hat{P}_{hh}(s,t) = \prod_{(s,t]} \left(1 + d\hat{A}_{hh}(u)\right) = \prod_{(s,t]} \left(1 - \frac{dN_{h\bullet}(u)}{Y_h(u)}\right), \quad h = 0,1, \tag{5.7}$$

with $dN_{0\bullet}(t) = \sum_{j=1}^k dN_{0j}(t)$ and $dN_{1\bullet}(t) = \sum_{j=2}^k dN_{1j}(t)$. Since the $\hat{A}_{hh}(\cdot)$ are continuous step-functions with a finite numbers of jumps, the product integral reduces to a simple product over the jump times on $(s,t]$. Note that the product is over times of any observed transition out of state h.

Estimation of $P_{01}(\cdot)$ in (C.1) is straightforward, since it is obtained by plug-in of the

estimators $\hat{P}_{00}$ and $\hat{P}_{11}$ given in (5.7). Similarly, estimators of the cumulative incidence probabilities in (5.2) and (5.1), as well as $S(\cdot)$ in (5.4), are given by the plug-in method. Note that in these equations the integrals reduce to sums, when the estimators are computed. For instance, $\hat{P}_{1h}(s,t)$ is obtained by summation over the transition times from 1 to h, that are the times when an event of type h is observed for a diseased patient.

5.3 Time-dependent covariates in the extended illness-death model for competing risks

Regression analysis on Markov multi-state models can be performed by the standard approach (Andersen and Keiding, 2002, Andersen et al., 1993, Chap. 7). Time-dependent covariates are also allowed.

Denote with $\boldsymbol{N}_i = (N_{0l,i}(t), l \in \{1,\ldots,k\}; N_{1h,i}(t), h \in \{2,\ldots,k\})$, for $i = 1,\ldots,n$ the multivariate counting process of individual i. Regression on covariates is made by specifying regression forms for $\alpha_{0l,i}(t; X_i(t))$ and $\alpha_{1h,i}(t; X_i(t))$, for $l \in \{1,\ldots,k\}$ and $h \in \{2,\ldots,k\}$, where $X_i(t)$ is the vector containing all the cause-specific covariates for individual i. Estimators of the baseline cumulative hazard functions can then be obtained by Breslow estimators, similarly to what was done for the application to breast cancer presented in Chapter 2.

When the interest is on predicting transition probabilities, the transition matrix is estimated as usual by the product integral

$$\hat{P}(s,t; X_0(t)) = \prod_{(s,t]} \left(\boldsymbol{I} + d\hat{\boldsymbol{A}}(u; X_0(t)) \right), \tag{5.8}$$

where $\hat{\boldsymbol{A}}(u; X_0(t))$ is the matrix of estimators of the integrated transition intensities. In order to obtain these estimators, the covariates need to be specified, as denoted by the given covariate vector $X_0(t)$.

A fundamental aspect so that the estimator in (5.8) is valid concerns the types of time-dependent covariates (Appendix B) included in the multi-state model. In case an internal time-dependent covariate is included in the regression analysis, predictions on the transition probabilities conditionally on given covariates, as in (5.8), are not possible.

We explain the reason of this by an informal example. Suppose the interest of a study is to predict some cumulative incidence probabilities under the extended illness-death

model for competing risks. Assume the regression models for the specific intensities $\alpha_{1h}(\cdot; X_I(\cdot))$, $h = 2, \ldots, k$, include an internal time-dependent covariate $X_I(\cdot)$. Then, if we wish to estimate some $P_{1h}(s,t)$, from (5.1) we observe that they are functions of $P_{11}(s, u-)$. This latter depends indirectly on $X_I(\cdot)$ through its dependence on all the cause-specific hazards $\alpha_{1h}(\cdot; X_I(\cdot))$, $h = 2, \ldots, k$ (as seen in equation (5.3)), but it also depends directly on the internal covariate, since $X_I(\cdot)$ carries information on the survival time of individuals. Therefore, given the observed covariate history up to time $u-$, $P_{11}(s, u-; X_I(u-))$ does not have anymore a meaningful interpretation. In fact, it is equal to one, and the consequence would be that all the cause-specific hazards α_{1h} are null.

5.3.1 Internal binary time-dependent covariates

In the previous subsection we illustrated the problems related to including an internal covariate into the regression model when predictions of the transition probabilities are of interest. In this context, the extended illness-death model for competing risks would provide a useful instrument when an internal binary time-dependent covariate would need to be studied. More specifically, when the binary time-dependent covariate is a simple one-step process, we might incorporate this process into the competing risks model. This means that the information given by the binary covariate is represented by the two additional states (0 and 1) of the extended competing risks model, presented in Section 5.2.

An example can be a study where it is important to take the binary covariate response/no response to a certain treatment into account. One may include this internal variable into a competing risks regression model, and thus investigate its effect on the cause-specific hazards, although cumulative incidence probabilities can not be estimated without specifying a model for the covariate. Otherwise, the extended competing risks model in Section 5.2, with transient states 0, 1 equal to, respectively, 'no response' and 'response', may be considered. Therefore, the probabilities of dying for a certain cause, for response and non-response patients, can be estimated and even compared. It may also be of interest to estimate the probability $P_{01}(0,t)$ to respond to the treatment within a certain time t.

A further interesting idea would consist of allowing patients to enter the study at time origin in either state 0 or 1. An initial distribution would need to be specified, i.e. the probabilities $\pi_0(0)$ and $\pi_1(0)$. This situation can be realistic when patients do not experience both the statuses described by states 0 and 1. For example, when studying

mortality due to different causes for some patients taking the role of HIV into account, one might be faced with children with or without HIV at birth, corresponding to entering in state 1 or 0, respectively.

Finally, we mention a possible extension of the model in Section 5.2 when the possibility of 'recovery' is also considered, that is when the transition from state 1 back to state 0 is possible, as in the illness-death model represented in Figure C.1. Explicit expressions for the transitions probabilities of this model can not be obtained anymore, although estimators based on the product integral representation can be computed, similarly to what was done in Section 5.2.

5.3.2 The time-dependent covariate 'duration in a state'

Within the extended competing risks model in Section 5.2, a very relevant aspect to study is the sojourn time into a certain state. Suppose we are interested in the duration in state 1 since the time T_1 of entrance in this state, and denote it with d. Therefore, the intensities regulating the transitions from 1 to the ending cause-specific events depend on the duration d and can be written as $\alpha_{1,h}(t,d)$, $h = 2,\ldots,k$, with $d = t - T_1$. We define $d = 0$ for $t < T_1$.

These intensities can be modelled by regression on the time-dependent covariate duration d, besides other covariates. An example is the Cox regression model, $\alpha_{1h}(t, t - T_1) = \alpha_{1h,0}(t)\exp\{(t - T_1)\beta\}$. Since the transition time T_1 is random, the duration $d = t - T_1$ is a simple random process, which is null until T_1, while after T_1 it increases linearly with time t. Let $\mathcal{F}_t$ and $\mathcal{X}_t$ be the filtrations generated by the observed multi-state process and the observed covariate d, respectively. Since T_1 is determined by the multi-state process itself, d is adapted to the filtration $\mathcal{F}_t$, and then the observed history of d is so that $\mathcal{X}_t \subset \mathcal{F}_t$. The covariate d can not be considered as determined in advance from time $t = 0$, since T_1 is unobserved at the time origin.

However, after entering state 1 at T_1, given the covariate history $\mathcal{X}_t$ with $t \geq T_1$, the duration d can be considered as determined. Thus, it may be thought of as a defined time-dependent covariate (Appendix B).

Since, d is assumed to influence the multi-state process only through the transition intensities $\alpha_{1h}(\cdot), h = 2,\ldots,k$, the transition probabilities from state 1 given the observed history $\mathcal{F}_{s-}$ are

$$P_{1h}(s,t;T_1) = \int_s^t P_{11}(s,u-;T_1)\alpha_{1h}(u,u-T_1)du, \qquad h = 2,\ldots,k, \quad T_1 < s \leq t,$$

where

$$P_{11}(s,u;T_1) = \exp\left(-\int_s^u \sum_{h=2}^{k} \alpha_{1h}(v, v-T_1)ds\right).$$

Similarly to equation (5.2), the cumulative incidence probabilities for $h = 2,\ldots,k$ can be expressed as follows

$$\begin{aligned} P_{0h}(s,t) &= \int_s^t \left[P_{00}(s,u-)\alpha_{0h}(u) + P_{01}(s,u-)\alpha_{1h}(u)\right] du \\ &= \int_s^t P_{00}(s,u-)\alpha_{0h}(u)du + \int_s^t P_{00}(s,u-)\alpha_{01}(u)P_{1h}(u,t|u)du, \end{aligned} \tag{5.9}$$

where

$$P_{1h}(u,t|u) = \int_u^t P_{11}(u, v-|u)\alpha_{1h}(v, v-u)dv,$$

and P_{00} is given by (5.3).

When transition intensities depend on $d = t - T_1$, which is studied as a time-dependent covariate, estimation of transition probabilities is straightforward, as seen just above, since in (5.9) the conditional probabilities $P_{1h}(u,t|u)$ are computed for all possible times u of transition to state 1. If we consider the extended illness-death model for competing risks where individuals are allowed to enter the study in either state 0 or 1 at the time origin, predictions on $\boldsymbol{P}$ can be possible only if the filtration $\mathcal{F}_0$ at time 0 contains information about the previous entry time T_1 in state 1. In this case, the duration can be considered as a completely predetermined time-dependent covariate. Otherwise, if information on T_1 is unknown for patients entering the study in state 1, then the sojourn time in this state can not be observed, and hence estimation under the model previously described can not be performed.

Interesting open questions arise in studying the extended competing risks model of Section 5.2 with duration dependence. Some complications, for instance, arise also when left-truncation is present, since information on T_1 might not be known for patients with delayed entries.

Discussion

The work in this thesis dealt with competing risks in survival analysis, both in the case of known specific causes and with the case of unknown (even if present) specific causes of the event. In the first case, we discussed the competing risks models and we focused on regression for the cumulative incidence probability. In the second case, where the event related to a certain group of diseased patients is recorded without any cause, regression models for relative survival were discussed. As shown by the present work, it is important to pay attention to inferential problems concerning dynamic aspects of models, such as time-dependent covariates and time-varying regression coefficients.

The competing risks setting was chosen as a very necessary statistical tool for studying the cardiotoxicity risk for patients with advanced breast cancer. Because of their severe status, it is known that these patients have a very high risk of dying, also during their chemotherapy treatment. That is why we can not neglect to consider the competing risk of dying for breast cancer even though the primary interest focuses on the risk of developing CHF. Ignoring the competing cause might lead to overlooking important features of the studied problem. Patients who died could potentially have developed CHF, but this event can never be observed.

The application of a competing risks analysis to the study of cardiotoxicity as a function of chemotherapy dosages led to very important new medical results. First of all, we found new recommended levels for the total dose administrated during Epirubicin chemotherapy, which were found to be lower than the one recommended in the literature (Ryberg et al., 1998). Moreover, the existing literature suggests a single level for all types of patients. We demonstrated that the optimal recommended dosage can vary substantially between groups of patients with different characteristics and risk factors.

In order to compute the optimal dosage levels corresponding to a 5% cardiotoxicity risk, we needed to treat cumulative dose as a time-dependent covariate. In handling time-dependent covariates, the implementation of the analysis was not trivial, since

the history of dose administration for each patient was needed, but statistical inference for the competing risks model did not require substantial modifications, since we were allowed to treat the time-dependent covariate as deterministic.

A drawback of the standard method used in Chapter 2 for regression analysis of competing risks data is that simple parameters, which explain directly the effects of covariates on the cause-specific cumulative incidence probabilities, are missing. The cumulative incidence probabilities are complex non-linear functions of the covariates and then, it is only possible to describe an indirect covariate effects only by predicting these probabilities for different given covariate patterns.

Problems about goodness-of-fit in case of a time-dependent covariate were investigated. Some of them were already pointed out by other authors (Therneau and Grambsch, 2000, Chap. 5), but we disagree on the usefulness of martingale residuals in suggesting possible correct functional form. Plots of martingale residuals both per-observation and per-subject might fail in investigating the functional form of a time-dependent covariate. We discussed about the need of cumulative martingale residuals (Lin et al., 1993) in model diagnostics, as they overcome problems related to time-dependency of covariates. A drawback of the type of residuals applied in Chapter 2 and the corresponding tests of hypotheses for each covariate, is that they are only valid if the Cox model is correct for all the remaining covariates (Scheike and Martinussen, 2004).

For relative survival, it was shown that the high flexibility of the additive nonparametric and semiparametric models, together with the inferential aspects described in Chapter 3, provides a very important alternative to the existing methods in this field, and on the other hand, a useful general extension of the more restrictive recent models.

The TRACE example demonstrates the need of new flexible survival models for modeling the excess hazards, which can deal with time-varying dynamics of covariates effects. In Chapter 3, we showed how the nonparametric and semiparametric versions of the additive excess hazard can easily handle these dynamics. We demonstrated when one or the other model is appropriate according to the responses of simulation based graphical and statistical tests about variation of effects over time. Even though inferential procedures described here are complicated in their expressions, when they concern finding equivalent asymptotic distributions of Gaussian processes, the great advantage is a very easy interpretation of results. In this connection, the statistical software, e.g. the R package `timereg` (Martinussen and Scheike, 2006, App. C) used in our application and presented in the Appendix, is an essential instrument.

In choosing between a proportional or an additive form for the excess hazards, problems about non-proportionality and large number of covariates under study should always be faced. An additional crucial problem, which was not studied in Chapter 3, concerns non-positive excess hazards in relative survival regression models. From a practical point of view, models about some situations as prevention studies, would need to allow the excess hazards to be negative, assuring however non-negative observed intensities. As pointed by Zahl (1996), the nonparametric additive excess hazards model overcomes this problem. A proportional excess hazards model can not be used in case of negative excess intensities, however, it is still possible to consider a possible excess intensity equal to zero (Sasieni, 1996).

Even though our suggestion for goodness-of-fit for relative survival presented in Chapter 4 is related to the recent interesting proposal by Stare et al. (2005), our approach has important advantages. First, our method does not need any critical choice of smoothing parameters (or parametric assumptions) for the baseline. Secondly, our procedure is asymptotically justified and will thus lead to asymptotically correct p-values and this is not true in general for the Stare et al. procedure.

Our suggestions about checking goodness-of-fit of the proportional excess hazards model and the additive excess models play an important role in a good model selection. An advantage of the supremum test described in Section 4.2 is that no specific deviations from proportionality need to be explicitly expressed. The drawback is however that the model is assumed to be correct with respect to all the other covariates when the proportionality assumption is investigated for a specific covariate. Nevertheless, this is a general problem faced also by the existing methods for goodness-of-fit of regression survival models.

In conclusion, presence of several (known or unknown) causes of an event of interest, typically death, are ubiquitous in biostatistics, and imply the necessity of studying problems in a competing risks or relative survival setting. Moreover, dynamic aspects are essential in providing a more accurate statistical description of the behaviour and effect of covariates in regression models. The results presented in the thesis contribute to illustrate these aspects, both from an applied point-of-view using real data, where we are faced with unexpected and realistic questions and complications, and by providing new theoretical improvements of the existing methodology.

Appendix A

R Code for Relative Survival Models

We show the basic R code concerning the application of the models presented in Section 3.4 to the TRACE data. The R package `timereg` can be downloaded at `http://staff.pubhealth.ku.dk/~ts/timereg.html`.

The dataset is called TR and it is structured with multiple observations for each patient in order to fulfil the conditions for studying time-dependent variables. The function `aalen.test` fits both the nonparametric and semiparametric additive excess hazards models presented in Sections 3.2 and 3.3. Commands for the former model are:

```
library(timereg);
dummy<-rnorm(nrow(TR));
fit1 <- aalen.test(Surv(start,stop,status>=7) ~ CHF+agec
+sex+diabetes+VF+const(dummy),data=TR,n.sim=300,max.time=6,
+offsets=TR$rate,id=TR$id,fix.gam=1);
summary(fit1)
```

In this example, the `Surv(start,stop,...)` setting is used for the time-dependent covariate `agec`, estimates are un-weighted and summary of the output shows the tests T_{1S} for non-significant effects and the tests T_{2S} and T_{2I} for time invariant effects. The offset `TR$rate` is the vector of expected mortality rates from the Danish population. The option `fix.gam` needs to be set equal to one in case of the nonparametric model. Further options are explained in the R help.

The following code,

```
plot.aalen(fit1,pointwise.ci=2,sim.ci=1)
```

provides graphics about the behaviour of the cumulative regression coefficients $B(t)$. The arguments `pointwise.ci` ≥ 1 and `sim.ci` ≥ 1 show, respectively, the 95% confidence intervals and the confidence bands based on 50 simulated processes under the null hypothesis.

The semiparametric additive excess hazard is given by:

```
fit2 <- aalen.test(Surv(start,stop,status>=7) ~ CHF+agec
+const(sex)+const(diabetes)+VF, data=TR,n.sim=300,
+max.time=6,offsets=TR$rate,id=TR$id);
summary(fit2);
plot.aalen(fit2,ylab="Test process",score=T)
```

The last plot, with the argument `score`, yields graphics about the observed processes used for computing T_{2S} and T_{2I} with 50 random realizations under the null hypothesis. Further options about `plot.aalen` are explained in the R help.

The function `pe.sasieni` fits the proportional excess hazards model described in Section 4.1 as follows:

```
fit3 <- pe.sasieni(Surv(start,stop,status>=7) ~ CHF+agec
+sex+diabetes+VF,data=TR,offsets=TR$rate,id=TR$id,
+max.time=6);
summary(fit3)
```

The summary provides statistics about the regression coefficients and tests for non-significant effects. The non-standardized version of the test for the hypothesis of proportionality of the hazards, based on cumulative martingale residuals and presented in Section 4.2, is also given in the summary.

Appendix B

Time-dependent Covariates

An overview of the different types of time-dependent covariates and their characteristics is described. Partial model specification and likelihood construction are reviewed for every type of such covariates and problems related to the survival function and predictions are illustrated.

B.1 Time-dependent covariates

Time-dependent covariates arise in regression models when the covariates change in time during the period of the study and their variation is influencing substantially the hazard functions. Let $\{X(t); 0 \leq t \leq \tau\}$ denote a time-dependent covariate process, where $[0, \tau]$ is the study period, and let $\mathcal{X}_t$ denote the filtration of the covariate history up to time t.

We assume to work under a right-censoring scheme and we suppose that a model for the hazard function $\alpha^\theta(t)$ is specified depending on a parameter θ. Let $\mathcal{F}'_t$ be the filtration generated by the observed survival data, as explained in the background Section 1.3.2.

When a regression model is specified, an extended history, $\mathcal{F}_t$, which incorporates also information about covariates needs to be considered. If regression is only on time-independent covariates, their information is expressed by the filtration $\mathcal{X}_0$ generated by all the covariates observed at the time origin 0. Hence, the extended observed history is $\mathcal{F}_t = \mathcal{F}'_t \vee \mathcal{X}_0$. If time-dependent covariates are included into the model, the observed filtration $\mathcal{F}_t$ contains also the covariate information up to time t, $\mathcal{X}_t$. It is then given by $\mathcal{F}_t = \mathcal{F}'_t \vee \mathcal{X}_t$, that is the smallest σ-algebra containing both the history generated

by the observed survival times and the covariate history $\mathcal{X}_t$ up to time t.

The first important assumption for regression modelling with a time-dependent covariate is that the covariate process is predictable with respect to $\mathcal{F}_t$, for instance being left continuous. The conditional hazard function given the history of the observed covariate process $X(t)$ is

$$\alpha^{\theta}(t; \mathcal{X}_t)dt = P\left\{t \leq T < t + dt | \mathcal{X}_t, T \geq t\right\}, \tag{B.1}$$

where T is a right-censored survival time. In some particular cases, the hazard function depends only on the current values of the covariates at time t.

Time-dependent covariates are divided into two general classes: External and internal covariates. In the literature they are sometimes also denoted as exogenous and endogenous covariates. A formal definition can be given for these two classes. A time-dependent covariate that satisfies the condition

$$P\left\{u \leq T < u + du | \mathcal{X}_u, T \geq u\right\} = P\left\{u \leq T < u + du | \mathcal{X}_t, T \geq u\right\} \tag{B.2}$$

for all u and t such that $u \leq t$, is called external (Kalbfleisch and Prentice, 2002). Hence, the hazard function at time u is influenced by the observed covariate history up to time u by the regression model, but the occurrence of a failure in $[u, u + du)$ is independent of the future path of the covariate after time u. This is equivalent to saying that a covariate is external if its future path up to any time t is not affected by the occurrence of a failure at time u. When the condition in (B.2) does not hold, a time-dependent covariate is called internal. The path of an internal covariate is affected by the occurrence of a failure time, since its existence depends on the survival of the individual. Therefore, its path carries information about the occurrence of a failure time.

Internal covariates are related to the random behavior of individuals under study, and consequently, they are observed only as long as individuals are at risk. In clinical biostatistics typical examples are disease complications, measurements recorded at the follow-up visits, such as biochemical and clinical characteristics, which give prognostic information on the status of patients. In general, external covariates have instead an observed path which is external to the individuals under study or it is not directly generated by their behavior in time. Some examples are the age of patients, levels of air pollution or the time since the disease diagnosis.

External covariates can be of two different types, as defined by Kalbfleisch and Prentice (2002). A time-dependent covariate is denoted as defined if it is deterministic in time, or if it varies in a predetermined way, since its path can be determined in advance. Age of patients and time since the disease diagnosis, if this latter is included in the information available at the time origin, are examples of that. A time-fixed covariate belongs to the class of defined covariates, since its value is given at the time origin and is constant for the duration of the study. The second type of external covariates is called ancillary. Their stochastic processes have distributions that do not involve the parameters of the regression model for survival times. An example of an ancillary time-dependent covariate is the measurement of air pollution used to predict the rate of asthma attacks.

B.2 Time-dependent covariates: Partial model specification and likelihood construction

The scope of this section is to remark the model specification and possible changes in the likelihood function due to regression on different types of time-dependent covariates.

As it was already described in Section 1.3.3, in case of right-censored data the full likelihood function for (θ, ϕ) factorizes as

$$L(\theta, \phi) = L_\tau^u(\theta, \phi) L_\tau^c(\theta), \tag{B.3}$$

with θ and ϕ being the parameter of interest and the nuisance parameter, respectively.

In regression models for the hazard, the first factor $L_\tau^u(\theta, \phi)$ may contain information about the additional parameter ϕ related to the distribution of the censoring mechanisms or/and the marginal distribution of covariates. The function $L_\tau^c(\theta)$ is the partial likelihood for θ and its form is given in equation (1.29). A model with such a factorization for the total likelihood can be partially specified, since computation of the partial likelihood for θ does not depend on the nuisance parameter ϕ and, generally, does not require specifying models for the covariates and the censoring mechanism.

When regression models include time-independent covariates, we are in the situation previously mentioned, and thus the partial likelihood can be written conditionally on the covariates, which are fixed given the filtration $\mathcal{X}_0$. However, this is not always true for time-dependent covariates, and complications arise when certain classes of

covariates are studied, as explained later in this section.

Suppose that $\mathcal{F}_t = \mathcal{F}'_t \vee \mathcal{X}_0$ is the observed filtration which incorporates only information on time-independent covariates. When a time-dependent covariate $X(t)$ is also included into the regression model, an important aspect to consider is whether its covariate process is adapted to $\mathcal{F}_t$ or not. It turns out that defined covariates are adapted to the filtration $\mathcal{F}_t$, and then, the history $\mathcal{X}_t$ generated by their observation up to time t is so that $\mathcal{X}_t \subset \mathcal{F}'_t \vee \mathcal{X}_0$ (Andersen et al., 1993, Chap. 3). This means that these covariates are either deterministic or their paths can be considered as being fixed in advance. Therefore, inference can be based on a partial likelihood which has the same form as $L^c_\tau(\theta)$ in (1.29).

Ancillary and internal covariates can instead be considered as random time-dependent covariates, and their process $X(t)$ is not adapted to the filtration $\mathcal{F}_t$. Therefore, the filtration needs to be extended so that $\mathcal{F}_t = \mathcal{F}'_t \vee \mathcal{X}_t$. However, when ancillary covariates are studied, inference on the parameter of interest θ can still be based only on the partial likelihood $L^c_\tau(\theta)$, conditioning on the observed paths. Since the ancillary covariate processes are completely external to the individuals under study, the model for these covariates does not depend on θ, and therefore does not need to be specified.

The main difference of internal covariates with respect to other covariate types is that they carry information about failure times of individuals. The hazard function has the same form as in (B.1), but now we can condition only on the covariate history $\mathcal{X}_{t-}$ up to the time just before t. Inclusion of internal covariates allows us to base inference on the partial likelihood $L^c_\tau(\theta)$, even though the full likelihood in (B.3) contains factors (included in $L^u_\tau(\theta, \phi)$) related to the marginal distribution of $X(t)$ given the history $\mathcal{X}_{t-}$ (Kalbfleisch and Prentice, 2002). Thus, we can avoid to specify a model for $X(t)$, but, in case this model depends on both the parameters θ and ϕ, inference based on the partial likelihood can be inefficient (Greenwood and Wefelmeyer, 1990).

B.3 Time-dependent covariates: Survival function and predictions

For external covariates, the survival function is well defined and, conditionally on the covariate history, it is given by $S(t; X(t)) = P\{T > t|\mathcal{X}_t\}$. Therefore it can be estimated without problems, given a certain covariate path up to time t.

For internal covariates the previous situation does not hold and care needs to be taken

in interpreting the survival function. Since the observation of $X(t)$ contains the failure information for an individual, the knowledge of its process up to time t would mean that the patient is still at risk at time t without having experienced the failure event, and therefore it would be $P\{T > t|\mathcal{X}_t\} = 1$. The conditional hazard in (B.1) can not be directly related to the survival function, as it is usually done (equation (1.2)), since in this case the survival distribution is meaningless and does not have any interpretation. The survival probability is not anymore a function only of the hazard function, but also of the random development of the covariates. Therefore, in order to provide an estimate for the survival probability, a distribution for the stochastic process of the internal covariate must be also specified.

When internal covariates are studied, predictions based on the model can not be made, because of the same reasons previously explained for the survival probability. The model is no longer partially specified and the parameters in the covariate model need also to be considered as parameters of interest.

Appendix C

The Illness-Death model

The illness-death model, also called the disability model, is a multi-state model for a nonhomogeneous process in continuous time with a finite number of states. This model is very useful in clinical studies where it is important to record whether or not, and how many times, the patient changes a certain clinical status before his/her possible death, and the aim is to study in time the rate of these events. For instance, before dying, patients may change status from being healthy to becoming diseased, and later on they may recover changing back into the healthy status.

We present this model, referring to the general formulas and notation about multi-state models given in Section 1.4.1. Its first formulation was discussed in the papers by Fix and Neyman (1951), Sverdrup (1965). A brief summary of the model is contained in the paper by Andersen and Keiding (2002) and several applications can be found in Andersen et al. (1993).

The state space is $\{0, 1, 2\}$, where 0 and 1 are transient states, representing the clinical statuses of individuals, and 2 is an absorbing state, corresponding to the ending event. Let us consider the probability space $(\Omega, \mathcal{Z}, P)$, where $\mathcal{Z}$ is the filtration generated by the stochastic process for the illness-death model. The time interval of interest is $\mathcal{T} = [0, \tau)$ with $\tau \leq \infty$. The stochastic development of the process is specified by the transition probabilities between states, $P_{hl}(s, t)$ with $h, l \in \{0, 1, 2\}$ and $s \leq t \in \mathcal{T}$, (Equation (1.33)) or, equivalently, by the matrix of transition intensities of the process, $\alpha_{hl}(t)$ with $h, l \in \{0, 1, 2\}$ (Equation (1.34)). The illness-death model and its possible transitions are illustrated in Figure C.1, where for clarity the states $0, 1$ and 2 are denoted by healthy, diseased and death, respectively. However, the states may of course represent other types of ending events or intermediate statuses.

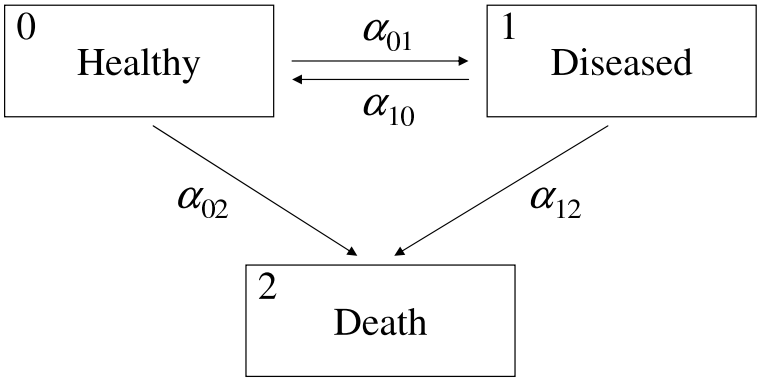

Figure C.1: *The illness-death model in the multi-state approach.*

In the present formulation we assume that all individuals are observed from the time origin, but this simple situation might not always be true and individuals might begin to be observed later (left-truncation). We suppose that all individuals are in state 0 at the time origin, i.e., the initial distribution of the process is $\pi_0(0) = 1$ and $\pi_h(0) = 0$ for $h = 1, 2$. Moreover, we restrict our description to the case of a Markov process underlying the illness-death model.

A simpler version of the illness-death model precludes the transition from the state 1 back to the initial state 0, which, for instance, means excluding the possibility of recovery for those individuals who are in the diseased status. The simpler version is given in Figure C.1, if the arrow from 1 to 0 is deleted, i.e., $\alpha_{10}(t) = 0$ for all $t \in \mathcal{T}$. Hereafter, the mathematical formulation is described for this latter version.

The transition probabilities for the simpler illness-death model are specified as follows:

$$P_{01}(s,t) = \int_s^t P_{00}(s,u-)\alpha_{01}(u)P_{11}(u,t)du, \quad s \leq t, \quad s,t \in \mathcal{T} \tag{C.1}$$

and

$$P_{02}(s,t) = \int_s^t P_{00}(s,u-)\alpha_{02}(u)du, \quad s \leq t, \quad s,t \in \mathcal{T}, \tag{C.2}$$

$$P_{12}(s,t) = \int_s^t P_{11}(s,u-)\alpha_{12}(u)du, \quad s \leq t, \quad s,t \in \mathcal{T}. \tag{C.3}$$

The probabilities of permanence in the states 0 and 1 have, respectively, the explicit

expressions

$$P_{00}(s,t) = \exp\left(-\int_s^t (\alpha_{02}(u) + \alpha_{01}(u))du\right), \tag{C.4}$$

$$P_{11}(s,t) = \exp\left(-\int_s^t \alpha_{12}(u)du\right), \tag{C.5}$$

since $\alpha_{00}(t) = -\sum_{h\neq l} \alpha_{hl}(t) = -(\alpha_{02}(t) + \alpha_{01}(t))$ and $\alpha_{11}(t) = -\alpha_{12}(t)$.

The marginal survival probability is given by $S(t) = P_{00}(0,t) + P_{01}(0,t)$, and it represents the probability of being alive, that is being either in state 0 or in state 1 at time t.

In some situations it may occur that the hazard of the transition from state 1 to state 2 depends on both the principal time scale t and the duration d of sojourn in state 1. One possible way of incorporating this dependence into the model is to consider the additional time scale d when studying the hazard. In this case, the transition intensities $P_{11}(s,t)$ and $P_{12}(s,t)$ are specified by replacing $\alpha_{12}(u)$ with $\alpha_{12}(u,d)$. Therefore, the instantaneous rate $\alpha_{12}(\cdot)$ depends on the random entry time into state 1. If this hazard depends on d only, then the illness-death model belongs to the class of semi-Markov models. Applications of such a model and some related theoretical aspects are proposed by Klein and Shu (2002).

Bibliography

Aalen, O. O. (1975). *Statistical Inference for a Family of Counting Processes*. PhD thesis, Department of Statistics, University of California, Berkeley.

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6:701–726.

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. In Klonecki, W., Kozek, A., and Rosinski, J., editors, *Lecture Notes in Statistics-2: Mathematical Statistics and Probability Theory*, pages 1–25. Springer-Verlag, New York.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8:907–925.

Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine*, 12:1535–1649.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5:141–150.

Andersen, P. K. (1998). Censored data. *Encyclopedia of Biostatistics*, 1:578–584.

Andersen, P. K., Abilstrom, S. Z., and Rosthøj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11:203–215.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11:91–115.

Andersen, P. K., Klein, J. P., and Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90:15–27.

Andersen, P. K. and Væth, M. (1989). Simple parametric and nonparametric models for excess and relative mortality. *Biometrics*, 45:523–535.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Bolard, P., Quantin, C., Abrahamowicz, M., Estève, J., Giorgi, R., Chadha-Boreham, H., Binquet, C., and Faivre, J. (2002). Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions: application to colon cancer. *Journal of Cancer Epidemiology and Prevention*, 7:113–122.

Bolard, P., Quantin, C., Estève, J., Faivre, J., and Abrahamowicz, M. (2001). Modelling time-dependent hazard ratios in relative survival: application to colon cancer. *Journal of Clinical Epidemiology*, 54:986–996.

Borgan, Ø. (1998). Aalen-Johansen estimator. *Encyclopedia of Biostatistics*, 1:578–584.

Brambilla, C., Rossi, A., Bonefante, V., Ferrari, L., Villiana, F., Crippa, F., and Bonadonna, G. (1986). Phase II study of doxorubicin versus epirubicin in advanced breast cancer. *Cancer Treatment Reports*, 70:261–266.

Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, 5:315–327.

Cortese, G. and Scheike, T. (2008). Dynamic regression models for relative survival. *Statistics in Medicine*. *In press*.

Courgeau, D. and Lelièvre, E. (1992). *Event History Analysis in Demography*. Clarendon, Oxford.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220.

De Gruttola, V. and Liao, Q. (1998). Truncated survival times. *Encyclopedia of Biostatistics*, 1.

Dickman, P. W., Sloggett, A., Hills, M., and Hakulinen, T. (2004). Regression models for relative survival. *Statistics in Medicine*, 23:51–64.

Ederer, F. and Heise, H. (1959). Instructions to IBM 650 programmers in processing survival computations. *Methodological note No. 10, End Results section, Bethesda, MD: National Cancer Institute*.

Efron, B. (1977). The efficiency Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565.

Estève, J., Benhamou, E., Croasdale, M., and Raymond, L. (1990). Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9:529–538.

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2:85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496–509.

Fisher, L. D. and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20:145–157.

Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 23:205–241.

Fleming, T. R. (1978a). Asymptotic distribution results in competing risks estimation. *Annals of Statistics*, 6:1071–1079.

Fleming, T. R. (1978b). Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *Annals of Statistics*, 6:1057–1070.

Fleming, T. R. and Harrington, D. P. (1993). *Counting Processes and Survival Analysis*. Wiley, New York.

Gandy, A., Therneau, T. M., and Aalen, O. O. (2007). Global tests in the additive hazards regression model. *Statistics in Medicine*, 27:831–844.

Giorgi, R., Abrahamowicz, M., Quantin, C., Bolard, P., Estève, J., Gouvernet, J., and Faivre, J. (2003). A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22:2767–2784.

Goldin, A., Vendini, J. M., and Geran, R. (1985). The effectiveness of of the anthracycline analog 4'-epidoxorubicin in the treatment of experimental tumours: A review. *Investigational New Drugs*, 3:3–21.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16:1141–1154.

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimator for partially specified filtered models. *Stochastic Processes and their Applications*, 36:353–370.

Greenwood, P. E. and Wefelmeyer, W. (1991). Efficient estimating equations for nonparametric filtered models. In Prabhu, N. U. and Basawa, I. V., editors, *Statistical Inference in Stochastic Processes*, pages 107–141. Marcel Dekker, New York.

Grønnesby, J. K. and Borgan, Ø. (1996). A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Analysis*, 2:315–328.

Hakulinen, T. (1982). Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics*, 38:933–942.

Hakulinen, T. and Tenkanen, L. (1987). Regression analysis of relative survival rates. *Applied Statistics*, 36:309–317.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480.

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis*, 5:239–264.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, 86:114–129.

Jensen, G. V. H., Torp-Pedersen, C., Hildebrandt, P., Kober, L., Nielsen, F. E., Melchior, T., Joen, T., and Andersen, P. K. (1997). Does in-hospital ventricular fibrillation affect prognosis after myocardial infarction? *European Heart Journal*, 18:919–924.

Jeong, J.-H. and Fine, J. (2006). Direct parametric inference for the cumulative incidence functions. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 55:187–200.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61:223–229.

Klein, J. P. and Shu, Y. (2002). Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research*, 11:117–139.

Kober, L., Torp-Pedersen, C., Carlsen, J., Bagger, H., Eliasen, P., and Lyngborg, K. (1995). A clinical trial of the angiotensin-converting-enzyme inhibitor trandolapril in patients with left ventricular dysfunction after myocardial infarction. Trandolapril Cardiac Evaluation (TRACE) Study Group. *New England Journal of Medicine*, 333:1670–1676.

Lambert, P. C., Smith, L. K., Jones, D. R., and Botha, J. L. (2005). Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine*, 24:3871–3885.

Lin, D. Y., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society Series B*, 62:711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80:557–572.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Annals of Statistics*, 23:1712–1734.

Martinussen, T. and Scheike, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika*, 89:283–298.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81:501–514.

Meira-Machado, L., de Una-Alvarez, J., Cardarso-Suarez, C., and Andersen, P. K. (2007). Multi-state models for the analysis of time to event data. *Research Report, Department of Biostatistics, University of Copenhagen*, 07/1:1–45.

Pohar, M. and Stare, J. (2006). Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81:272–278.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Geb.*, 51:269–286.

Ryberg, M., Nielsen, D., Cortese, G., Nielsen, G., Andersen, P. K., and Skovsgaard, T. (2008). New insight in epirubicin cardiac toxicity. competing risks analysis of 1097 breast cancer patients. *Journal of National Cancer Institute.*, page *In revision*.

Ryberg, M., Nielsen, D., Skovsgaard, T., Hansen, J., Jensen, B. V., and Dombernowsky, P. (1998). Epirubicin cardiotoxicity: An analysis of 469 patients with metastatic breast cancer. *Journal of Clinical Oncology*, 16:3502–3508.

Sasieni, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In Klein, J. and Goel, P., editors, *Survival Analysis: State of the Art*, pages 249–263. Kluwer, Dordrecht.

Sasieni, P. D. (1996). Proportional excess hazards. *Biometrika*, 83:127–141.

Scheike, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Analysis*, 8:247–262.

Scheike, T. H. (2004). Time-varying effects in survival analysis. In Balakrishnan, N. and Rao, C. R., editors, *Handbook of Statistics 23*, pages 61–85. Elsevier B.V., North Holland.

Scheike, T. H. and Martinussen, T. (2004). On efficient estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics*, 31:51–62.

Scheike, T. H. and Zhang, M.-J. (2002). An additive-multiplicative Cox-Aalen model. *Scandinavian Journal of Statistics*, 28:1328–1355.

Scheike, T. H. and Zhang, M.-J. (2004). Predicting cumulative incidence probability: Marginal and cause-specific modelling. *Research Report, Department of Biostatistics, University of Copenhagen*, 04/3.

Scheike, T. H. and Zhang, M.-J. (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*, 34:17–32.

Scheike, T. H., Zhang, M.-J., and Gerds, T. A. (2007). Predicting cumulative incidence probability by direct binomial regession. *Biometrika*. *In Press*.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69:239–241.

Sidney, I. (1992). *Adventures in Stochastic Processes*. Birkhäuser, Boston.

Stare, J., Pohar, M., and Henderson, R. (2005). Goodness of fit of relative survival models. *Statistics in Medicine*, 24:3911–3925.

Sverdrup, E. (1965). Estimates and tests procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skandinavisk Aktuarietidsskrift*, 48:184–211.

Swain, S. M., Whaley, S. F., and Ewer, S. M. (2003). Congestive heart failure in patients treated with doxorubicin. *Cancer*, 97:2869–2879.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale based residuals for survival models. *Biometrika*, 77:147–160.

Tormeys, D. C. (1975). Adriamycin (NSC-123 127) in breast cancer: An overview of studies. *Cancer Chemotherapy Reports*, 6:319–327.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences in USA*, 72:20–22.

Tsiatis, A. A. (1998). Competing risks. *Encyclopedia of Biostatistics*, 1:5–10.

Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *Journal of the American Statistical Association*, 79:649–652.

Zahl, P.-H. (1995). A proportional regression model for 20 year survival of colon cancer in Norway. *Statistics in Medicine*, 14:1249–1261.

Zahl, P.-H. (1996). A linear non-parametric regression model for the excess intensity. *Scandinavian Journal of Statistics*, 23:353–364.

Zahl, P.-H. and Tretli, S. (1997). Long-term survival of breast cancer in Norway by age and clinical stage. *Statistics in Medicine*, 16:1435–1449.